KTH VETENSKAP OCH KONST

Royal Institute of Technology

MACHINE LEARNING 2 – DGM, CH 8

Lecture JL1

Royal Institute of Technology

Computational Biology

Machine Learning – a main tool

Jens Lagergren

THIS LECTURE

- Probability?
- DGM
- Basic definitions
- Examples
- Learning parameters - given complete data
- Illustrating a known model

PRODUCT RULE: CONDITIONING

$$p(x, y) = p(y)p(x|y)$$

or

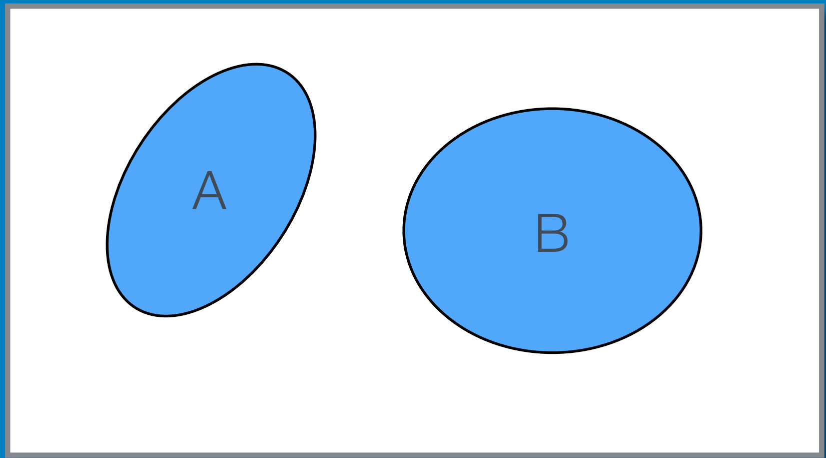
$$p(x|y) = \frac{p(x, y)}{p(y)}$$

BAYES RULE

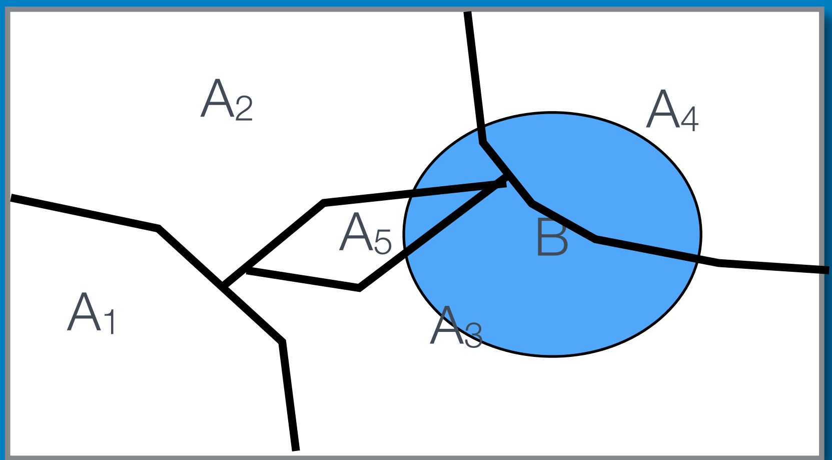
$$p(X|Y) = \frac{p(X,Y)}{p(Y)} = \frac{p(X)p(Y|X)}{\sum_x p(x)p(Y|x)}$$

SUM RULE: EXCLUSIVE & EXHAUSTIVE

Exclusive



Exclusive & exhaustive



- Exclusive

$$p(A \text{ or } B) = p(A) + P(B)$$

- Exclusive & exhaustive

$$p(B) = \sum_i p(B, A_i) = \sum_i p(A_i) p(B|A_i)$$

BERNOULLI AND CATEGORICAL

$$\mathrm{Ber}(x|\theta) = \begin{cases} \theta & \text{if } x = 1 \\ 1 - \theta & \text{if } x = 0 \end{cases}$$

$$\mathrm{Cat}(x|\boldsymbol{\theta}) = \theta_x$$

- ★ One or several (unordered) coin tosses
- ★ A dice (possibly biased)

MLE FOR CATEGORICAL

- Likelihood $$p(D) = \prod_{i \in [k]} \theta_i^{N_i}$$
- where $$\sum_{i \in [k]} \theta_i = 1$$
- as well as loglikelihood $$p(D) = \sum_{i \in [k]} N_i log \theta_i$$
- is maximized by

MLE FOR CATEGORICAL

- Likelihood $$p(D) = \prod_{i \in [k]} \theta_i^{N_i}$$
- where $$\sum_{i \in [k]} \theta_i = 1$$
- as well as loglikelihood $$p(D) = \sum_{i \in [k]} N_i log \theta_i$$
- is maximized by $$\theta_i = \frac{N_i}{\sum_{i \in [k]} N_i}$$

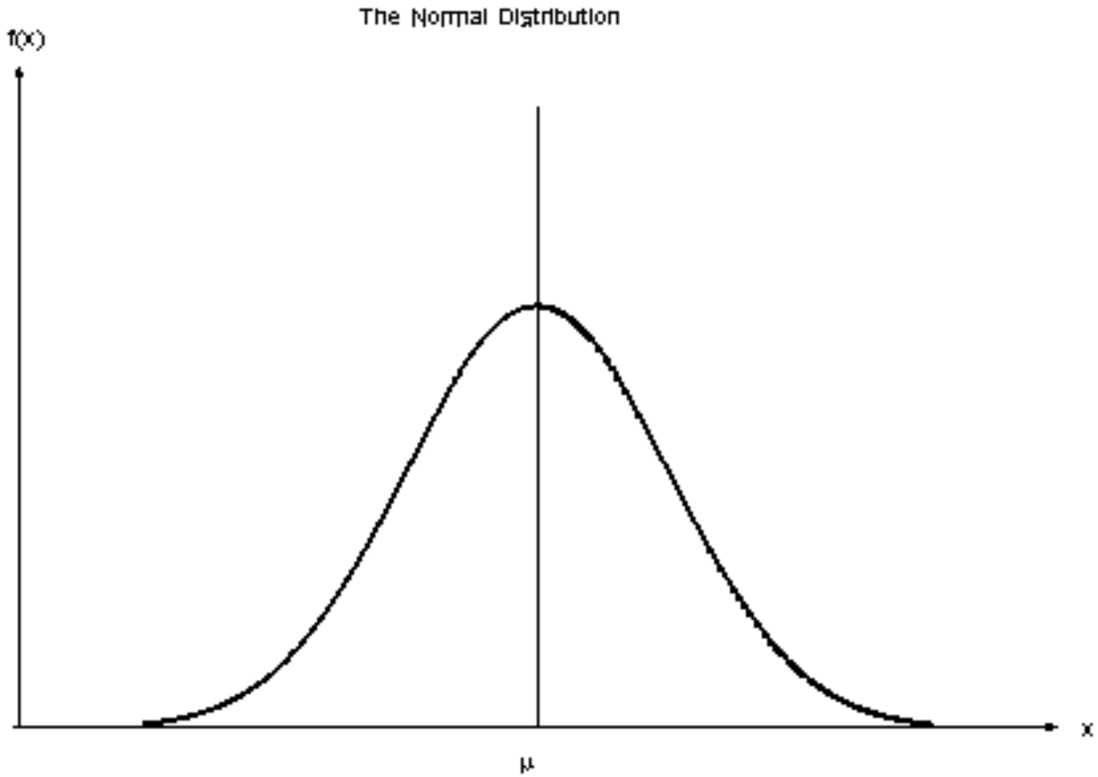
A TABLE

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010

VISUAL ACCESSIBILITY

Table A-1 The Standard Normal Distribution

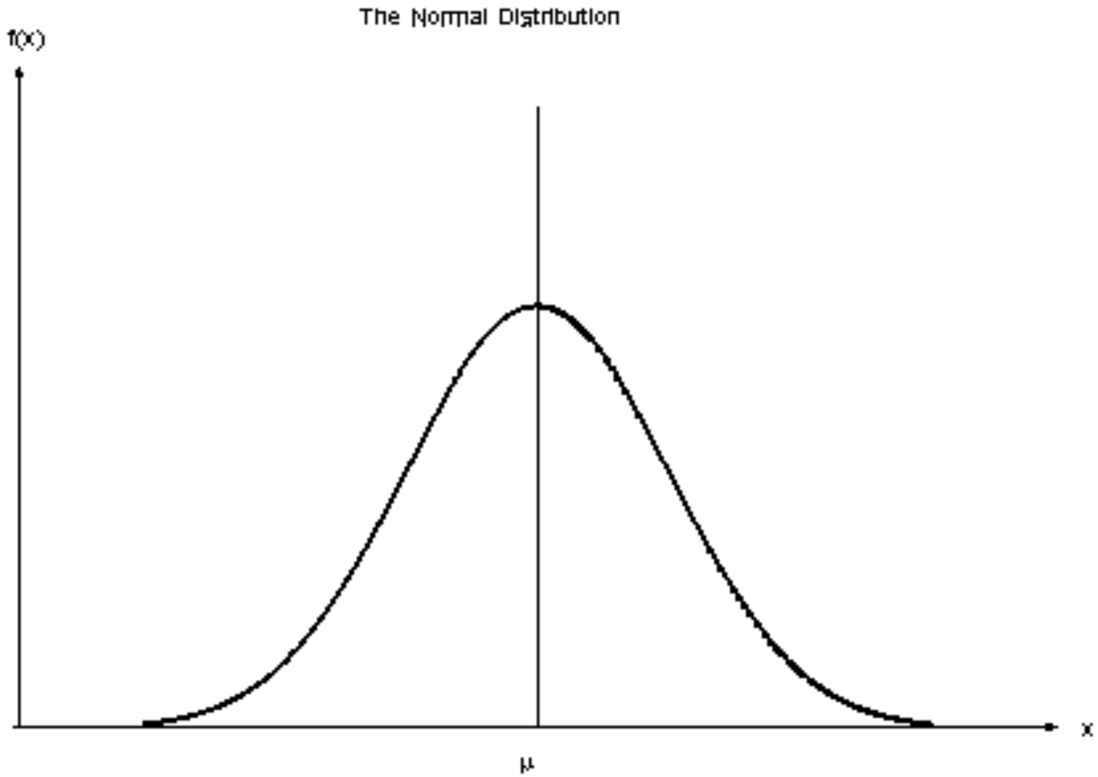
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010



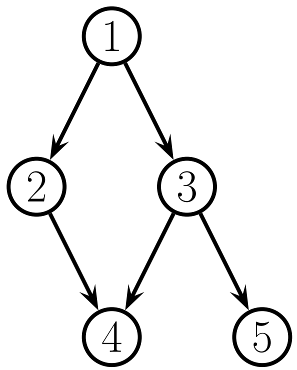
MATHEMATICAL TREATMENT

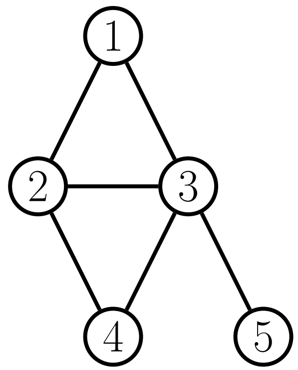
Table A-1 The Standard Normal Distribution

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010



$$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$





Directed graphical model

- DAG
- vertices r.v.s
- equipped with local CPDs
- allows causal like dependencies

Undirected graphical model - Markov Random Fields

- graph
- vertices r.v.s
- equipped with local “factors”

GRAPHICAL MODELS

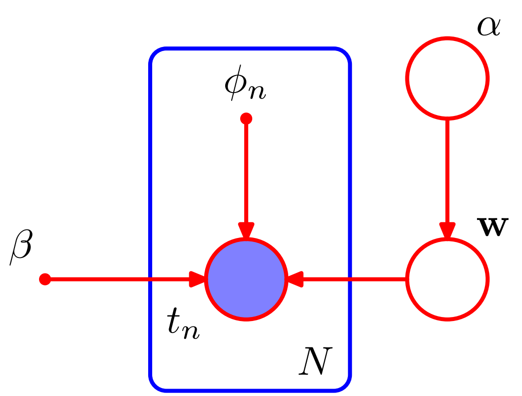
REPRESENTING AND WORKING WITH DISTRIBUTIONS

- For all but the smallest n, the explicit representation of the joint distribution is *unmanageable from every perspective*.
  - Computationally, it is very *expensive to manipulate* and generally *too large to store* in memory.
  - Cognitively, it is *impossible to acquire so many numbers* from a human expert; moreover, the numbers are very small and *do not correspond to events that people can reasonably contemplate*.
  - Statistically, if we want to learn the distribution from data, we would *need ridiculously large amounts of data to estimate* this many parameters robustly.
- These problems were the *main barrier* to the adoption of probabilistic methods for expert systems *until the development of the methodologies we now will consider.*

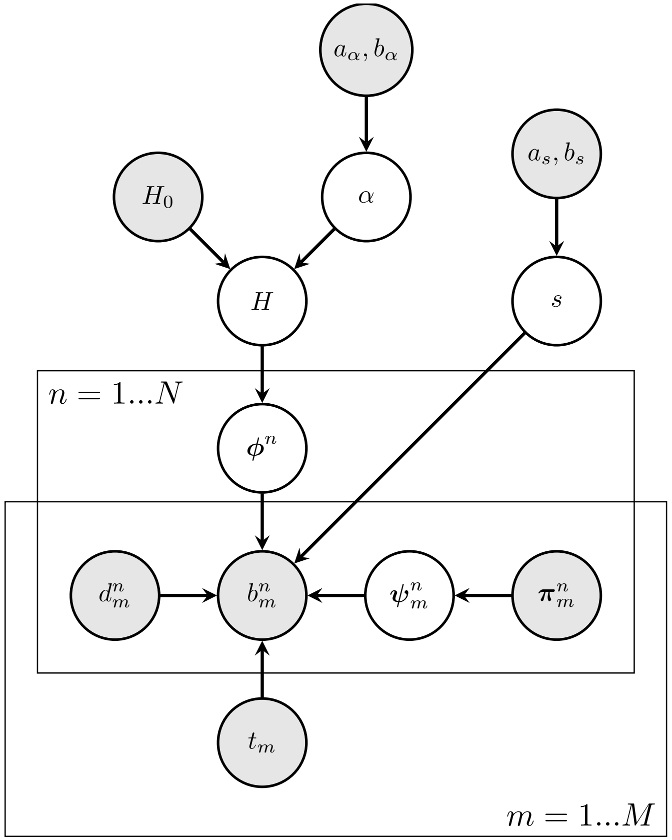
VISUALIZATION

- Another application
  - describe and visualize a “designed model” or a distribution and, in particular, its dependencies

Figure 10.8 Probabilistic graphical model representing the joint distribution (10.90) for the Bayesian linear regression model.

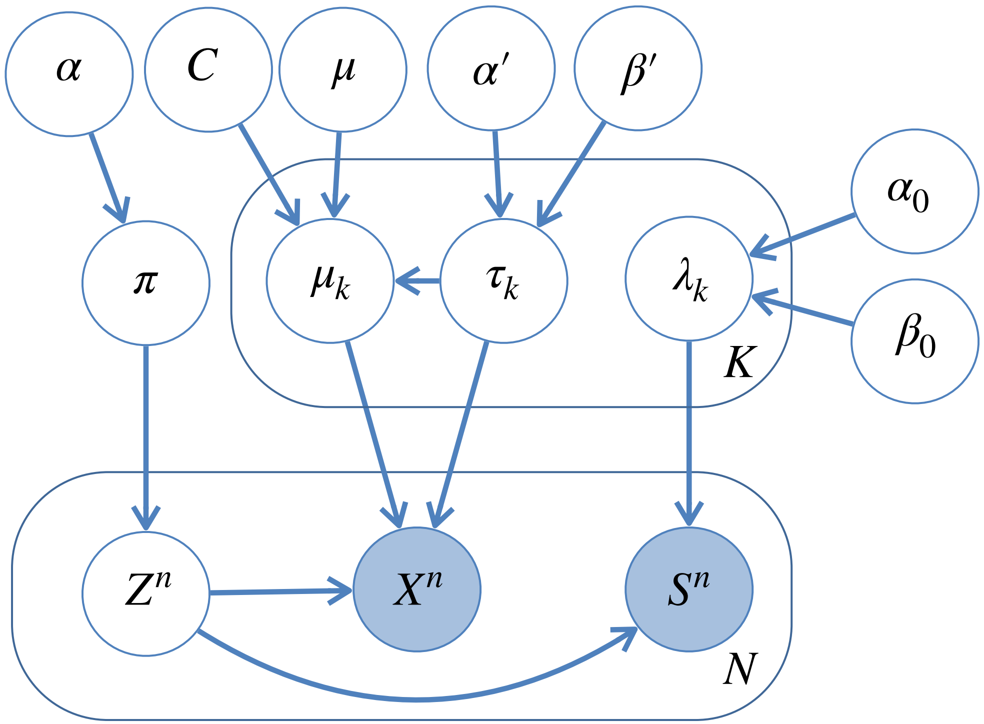


PYCLONE

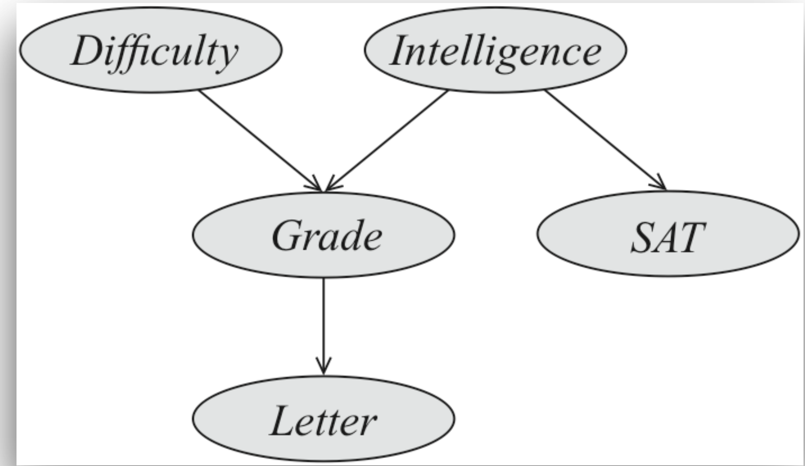


$$
\begin{aligned}
\alpha &\sim \mathrm{Gamma}(a_\alpha, b_\alpha) \\
H_0 &= \mathrm{Uniform}([0,1]^M) \\
H|\alpha, H_0 &\sim \mathrm{DP}(\alpha, H_0) \\
\boldsymbol{\phi}^n|H &\sim H \\
\boldsymbol{\psi}^n_m|\boldsymbol{\pi}^n_m &\sim \mathrm{Categorical}(\boldsymbol{\pi}^n_m) \\
\boldsymbol{\psi}^n_m &= (g^n_{m,\mathrm{N}}, g^n_{m,\mathrm{R}}, g^n_{m,\mathrm{V}}) \\
&\text{either} \\
b^n_m|d^n_m, \boldsymbol{\psi}^n_m, \phi^n_m, t_m &\sim \mathrm{Binomial}(d^n_m, \xi(\boldsymbol{\psi}^n_m, \phi^n_m, t_m)) \\
&\text{or} \\
s|a, b &\sim \mathrm{Gamma}(a_s, b_s) \\
b^n_m|d^n_m, \boldsymbol{\psi}^n_m, \phi^n_m, t_m, s &\sim \mathrm{BetaBinomial}(d^n_m, \xi(\boldsymbol{\psi}^n_m, \phi^n_m, t_m), s) \\
&\text{where} \\
\xi(\boldsymbol{\psi}, \phi, t) &= \tfrac{(1-t)c(g_\mathrm{N})}{Z}\mu(g_\mathrm{N}) + \tfrac{t(1-\phi)c(g_\mathrm{R})}{Z}\mu(g_\mathrm{R}) + \\
&\quad \tfrac{t\phi c(g_\mathrm{V})}{Z}\mu(g_\mathrm{V}) \\
Z &= (1-t)c(g_\mathrm{N}) + t(1-\phi)c(g_\mathrm{R}) + \\
&\quad t\phi c(g_\mathrm{V})
\end{aligned}
$$

THE SUPER-EPI GM



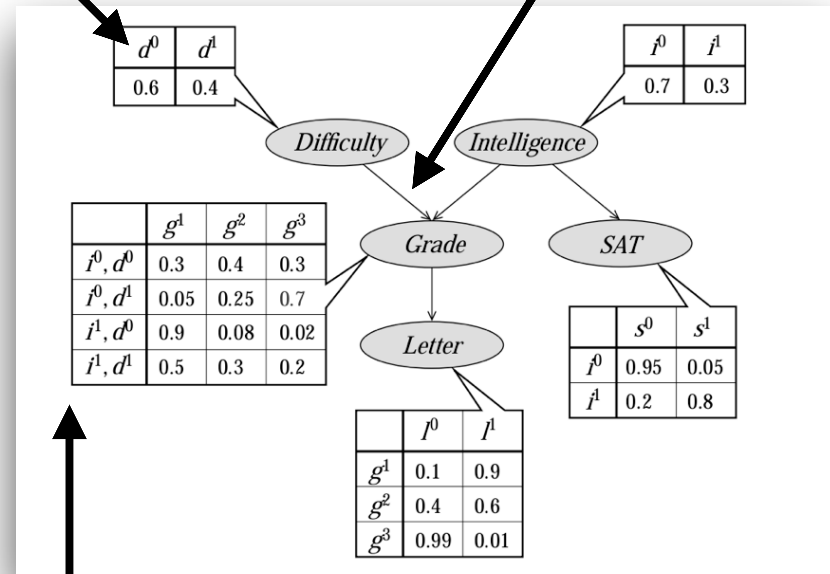
DGM - GRAPH AND CPDS VS JOINT



P(D,I,G,S,L)

d has value 0

DAG



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

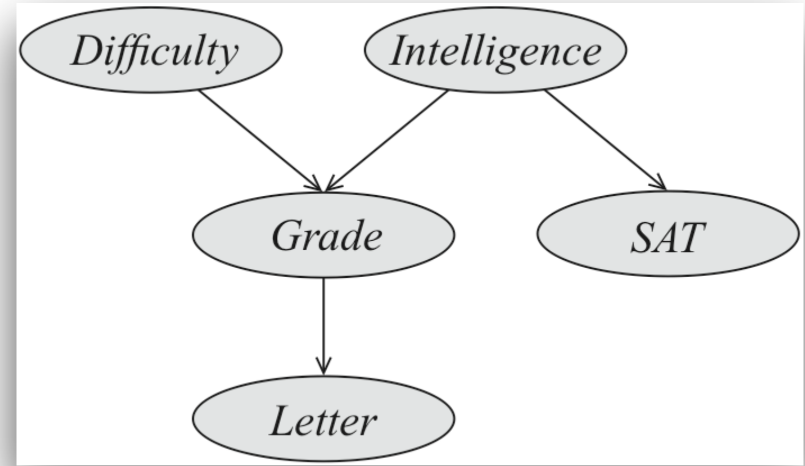
	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

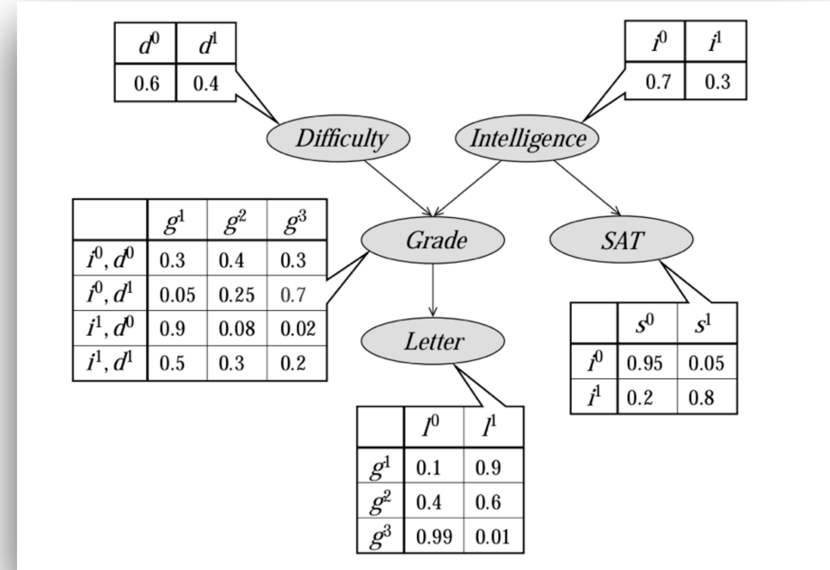
	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

CPD (Conditional Prob. Dist.)

- CPD - table, i.e., categorical
- Gaussian

THREE LEVELS



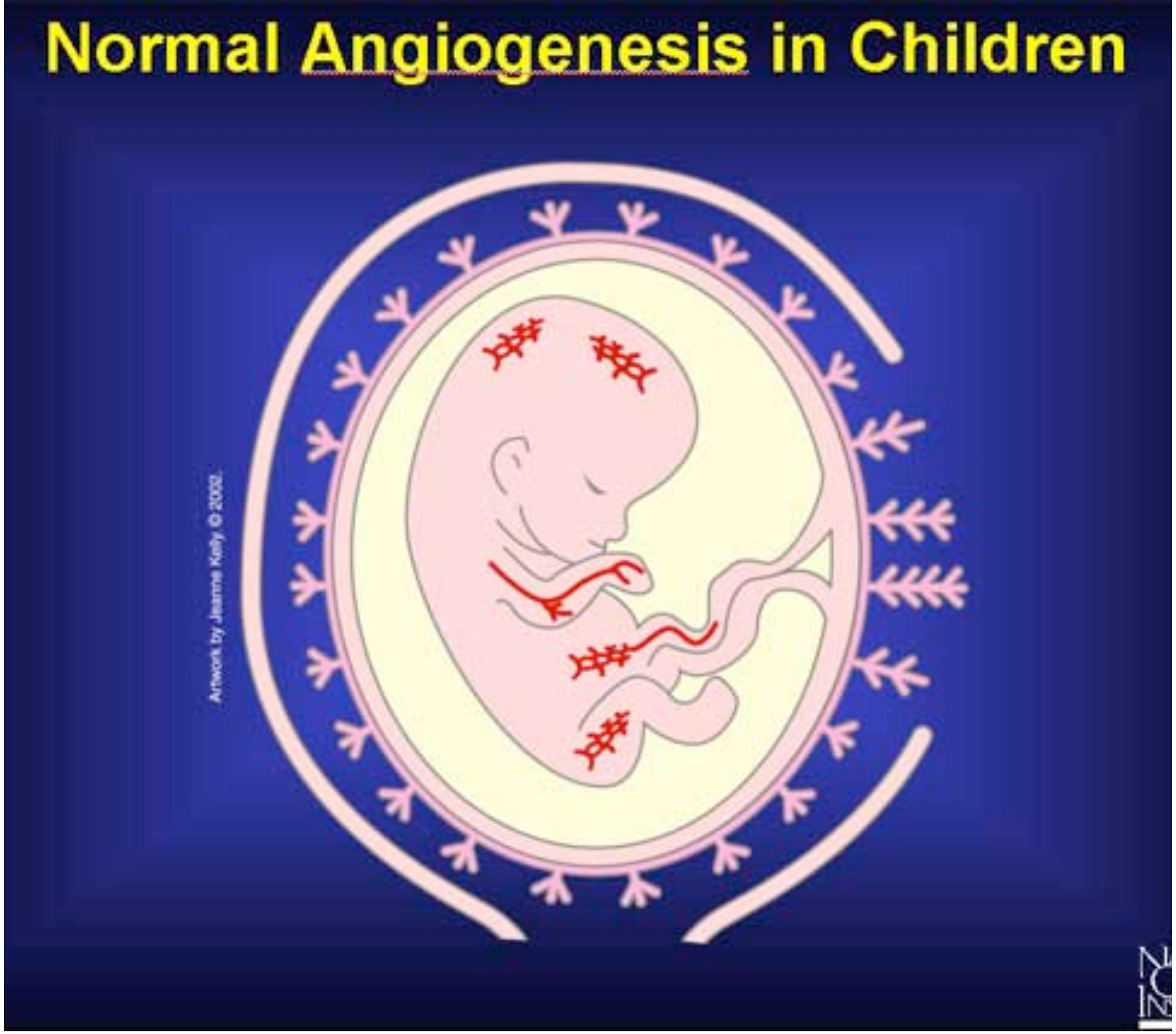


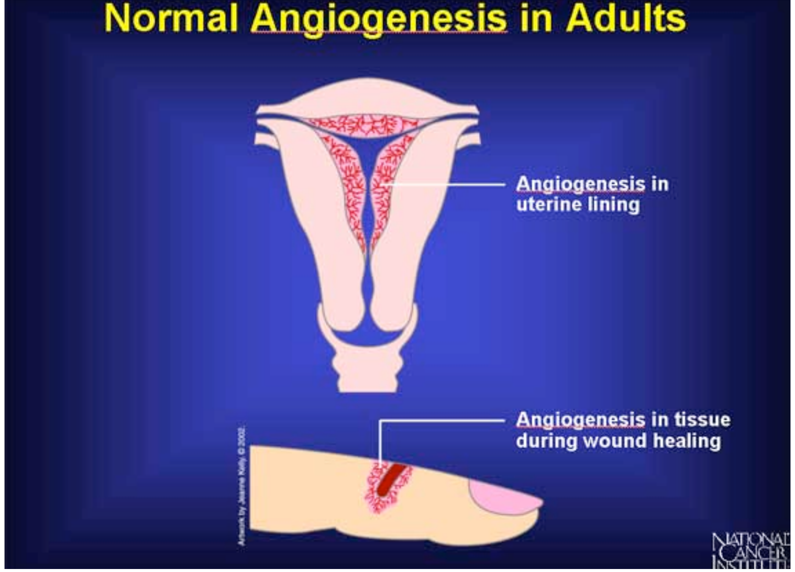
- Inference: given G and θ, compute probabilities
- or marginalize
- Parameter learning: given G and D, learn θ
- Structure learning: given D, learn G and θ

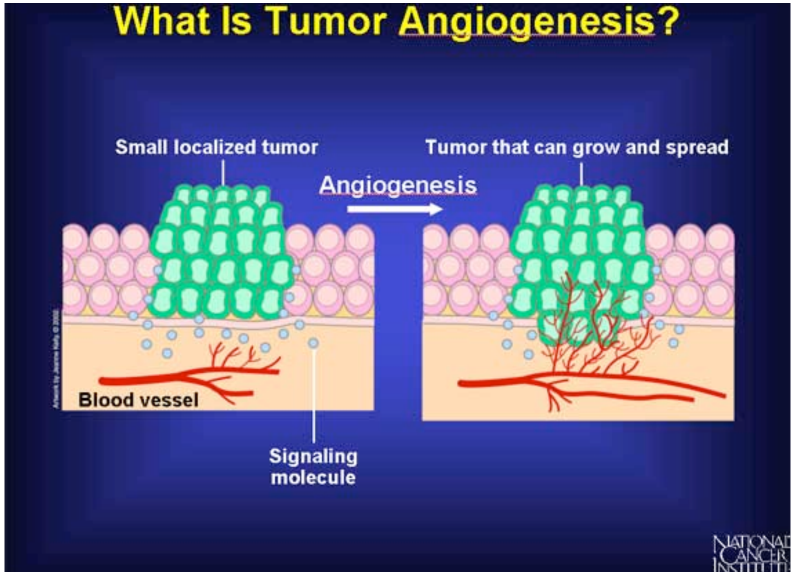
Marginalize often hard

Easy for observable data

Hard unless tree-like, doable in practice for observable







ABERRATION DEPENDENCIES - EX. ANGIOGENESIS

SOMATIC EVOLUTION

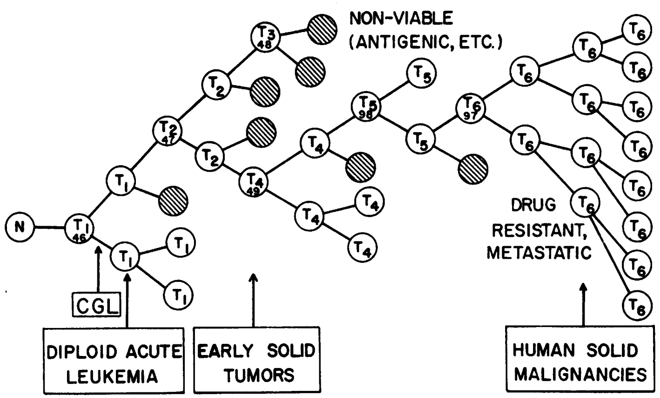
The Clonal Evolution of Tumor Cell Populations

Acquired genetic lability permits stepwise selection of variant sublines and underlies tumor progression.

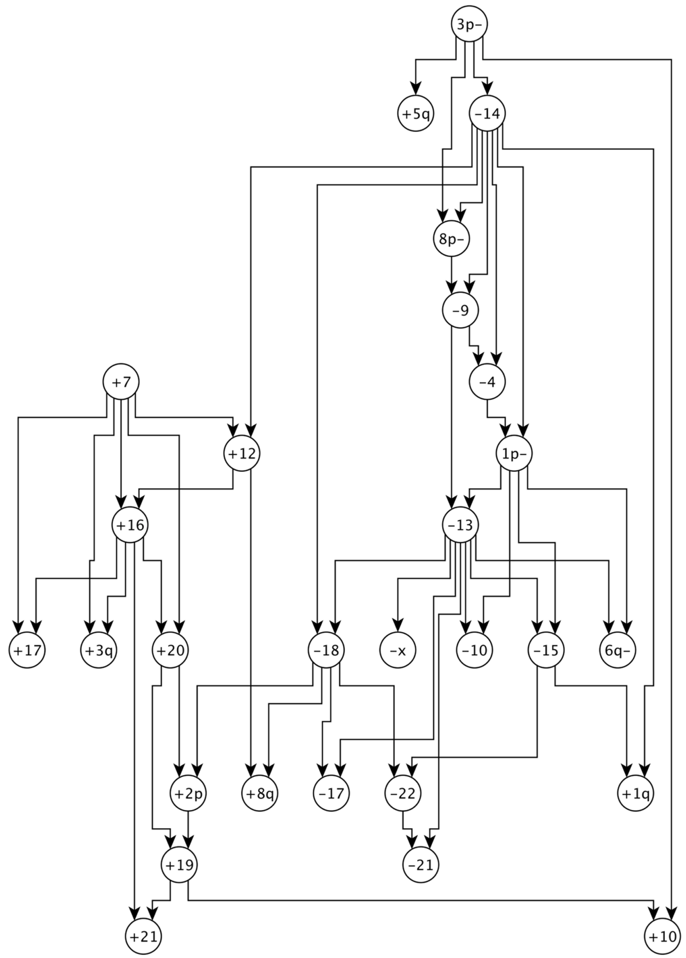
Peter C. Nowell

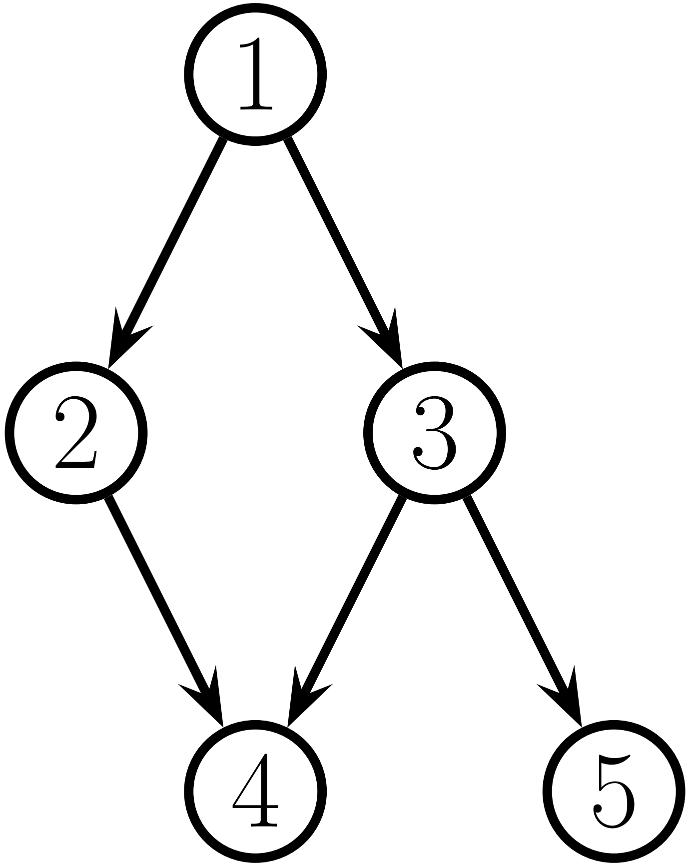
The author is professor of pathology, School of Medicine, University of Pennsylvania, Philadelphia 19174.

1 OCTOBER 1976 SCIENCE, VOL. 194



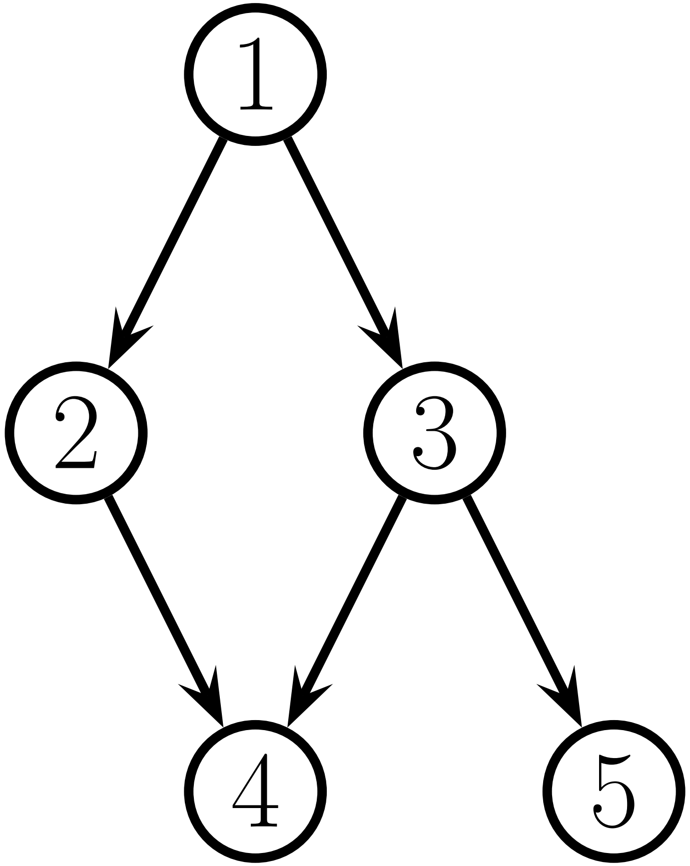
Oncogenetic network





TERMINOLOGY

- ★ Parent
- ★ Child
- ★ Family
- ★ Root
- ★ Leaf
- ★Neighbors



TERMINOLOGY

- ★ Degree (in and out)
- ★ Path (directed or not)
- ★ Cycle (directed or not)
- ★ Directed Acyclic Graph (DAG)
- ★ Topological order (parents < child)
- ★Ancestors

CPD - BERNOULLI OR CATEGORICAL

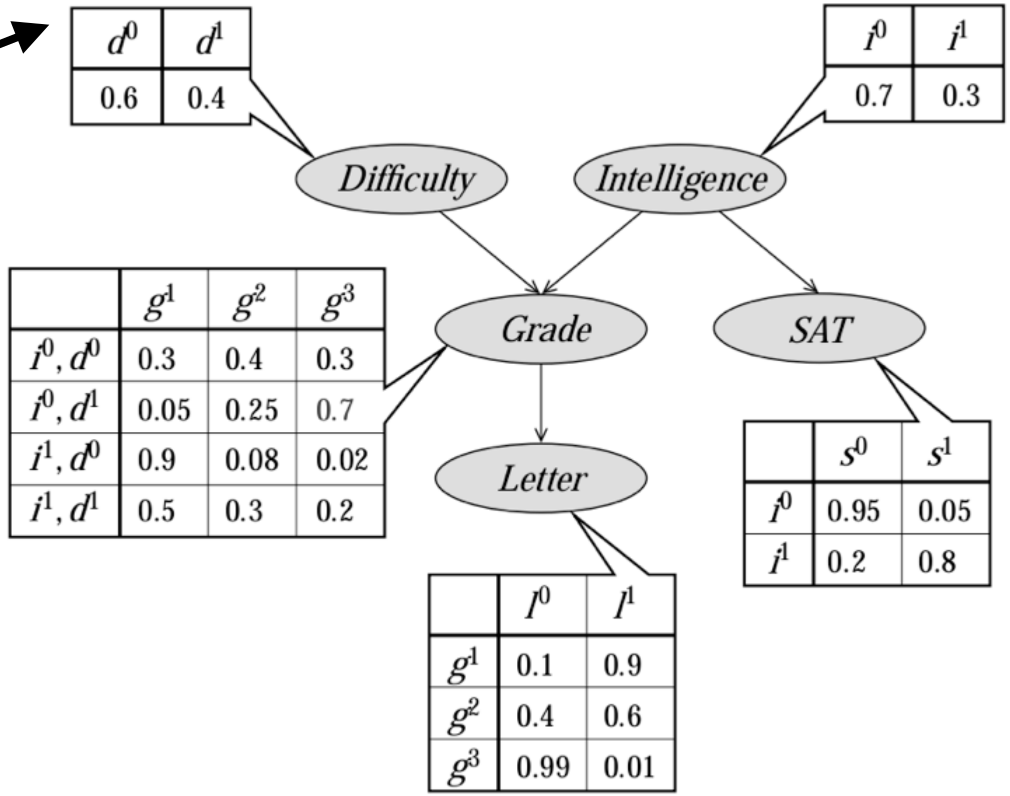
$$\mathrm{Ber}(x|\theta) = \begin{cases} \theta & \text{if } x = 1 \\ 1 - \theta & \text{if } x = 0 \end{cases}$$

$$\mathrm{Cat}(x|\boldsymbol{\theta}) = \theta_x$$

- ★ One or several (unordered) coin tosses
- ★ A dice (usually biased)

NOTATION

d has value 0



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

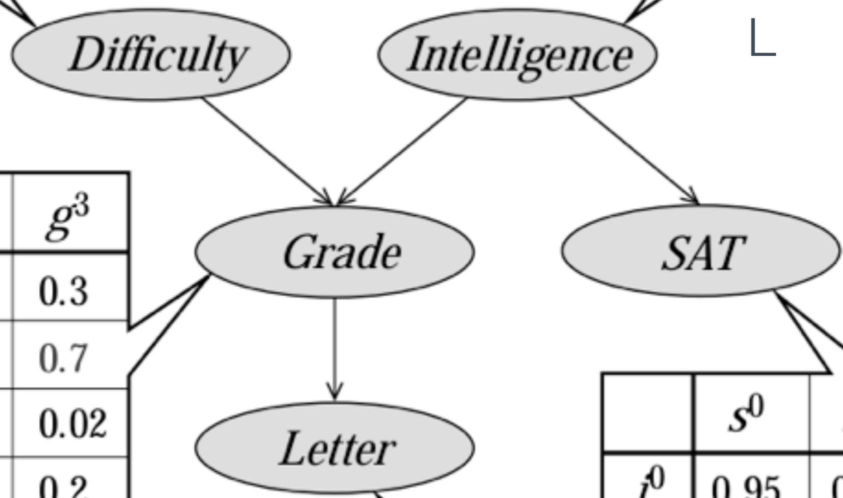
STUDENT EXAMPLE

d^0	d^1
0.6	0.4

L H

i^0	i^1
0.7	0.3

L H



	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

B L

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

L H

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

L B

B - better

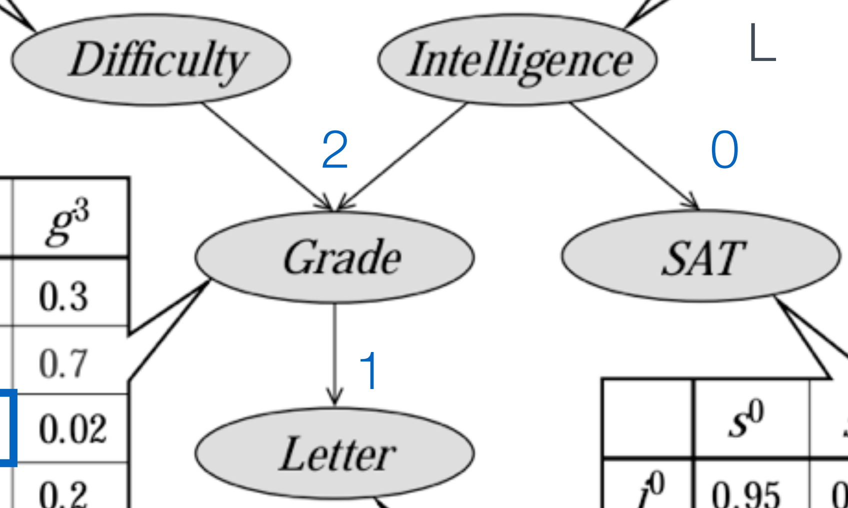
H - higher

L - less

STUDENT EXAMPLE

d^0	d^1
0.6	0.4
L	H

i^0	i^1
0.7	0.3
L	H



	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2
	B		L

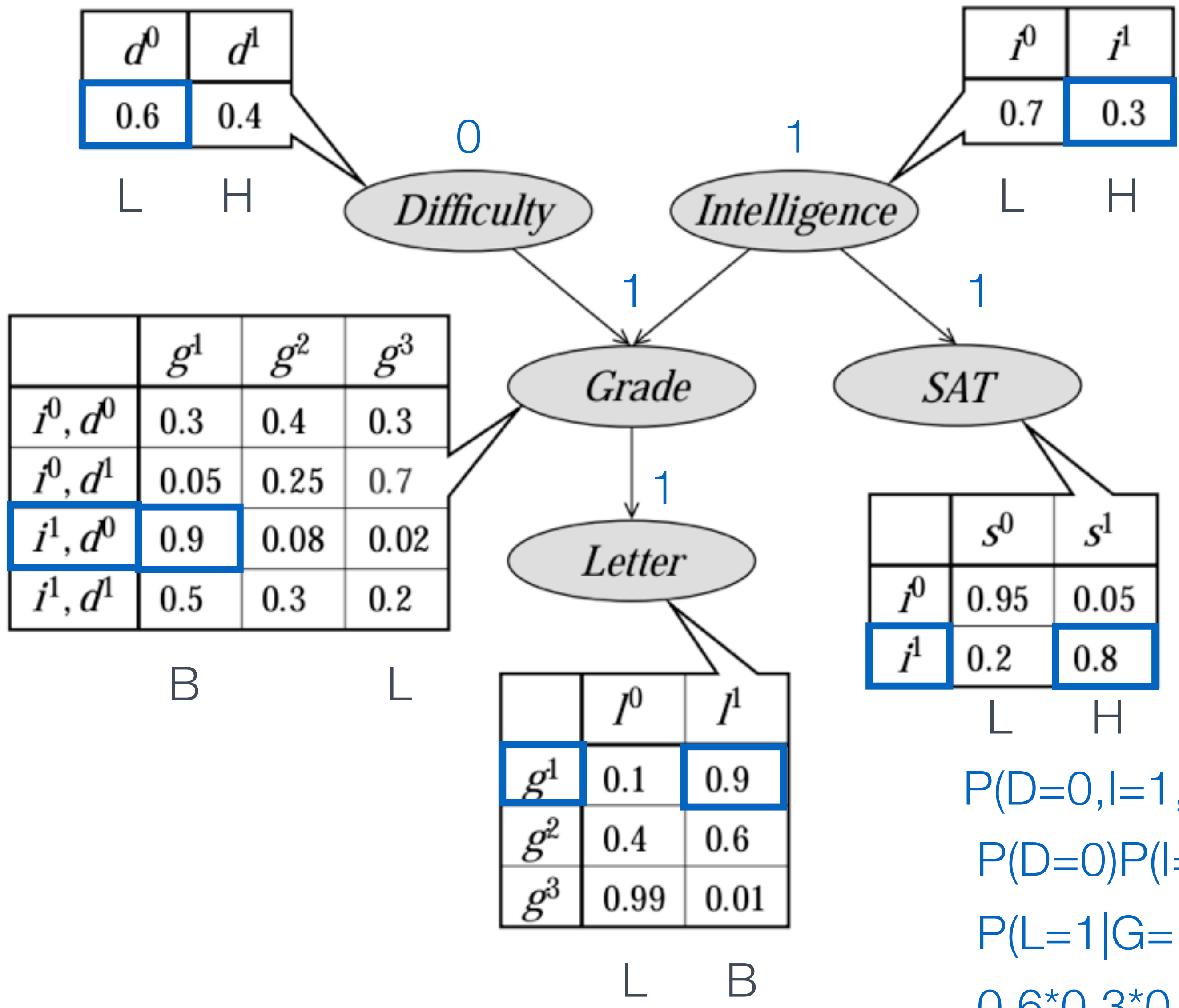
	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8
	L	H

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01
	L	B

P(D=0,I=1,G=2,L=1,S=0) =
P(D=0)P(I=1)P(G=2|D=0,I=1)
P(L=1|G=2)P(S=0|I=1)=
0.6*0.3*0.08*0.6*0.2

B - better
H - higher
L - less

STUDENT EXAMPLE



d^0	d^1
$\boxed{0.6}$	0.4

i^0	i^1
0.7	$\boxed{0.3}$

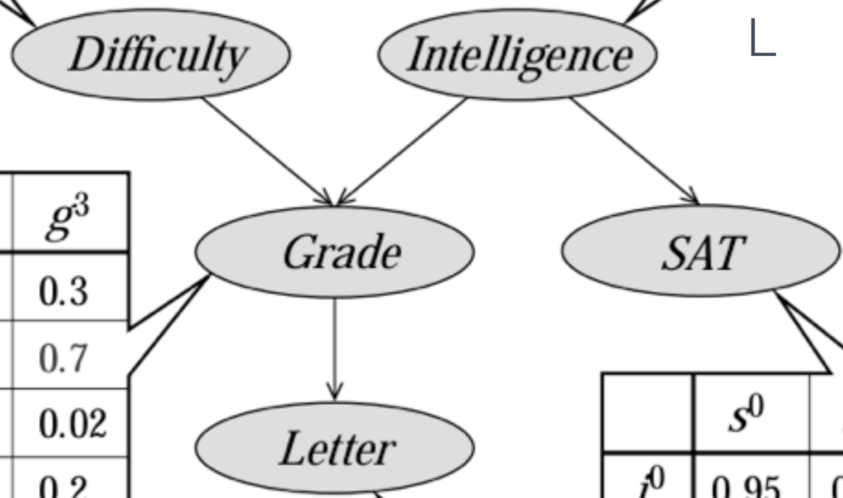
	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
$\boxed{i^1, d^0}$	$\boxed{0.9}$	0.08	0.02
i^1, d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
$\boxed{i^1}$	0.2	$\boxed{0.8}$

	l^0	l^1
$\boxed{g^1}$	0.1	$\boxed{0.9}$
g^2	0.4	0.6
g^3	0.99	0.01

P(D=0,I=1,G=1,L=1,S=1) =
P(D=0)P(I=1)P(G=1|D=0,I=1)
P(L=1|G=1)P(S=1|I=1)=
0.6*0.3*0.9*0.9*0.8

STUDENT EXAMPLE



d^0	d^1
0.6	0.4

L H

i^0	i^1
0.7	0.3

L H

	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

B L

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

L H

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

L B

P(D=0,I=1,G=2,L=1,S=0) =
P(D=0)P(I=1)P(G=2|D=0,I=1)
P(L=1|G=2)P(S=0|I=1)=
0.6*0.3*0.08*0.6*0.2

P(D=0,I=1,G=1,L=1,S=1) =
P(D=0)P(I=1)P(G=1|D=0,I=1)
P(L=1|G=1)P(S=1|I=1)=
0.6*0.3*0.9*0.9*0.8

FACTORIZATION OVER G

$$p(x_1, \ldots, x_N) = \prod_{n=1}^{N} p(x_n | \boldsymbol{x}_{\mathrm{pa}(x_n)})$$

p can be factorized over G if it can be expressed as above

INFERENCE – THE CHAIN RULE

$$p(\underbrace{\boldsymbol{x}_{[V]}}_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_V}) = p(\boldsymbol{x}_1)p(\boldsymbol{x}_2|\boldsymbol{x}_1)p(\boldsymbol{x}_3|\boldsymbol{x}_1,\boldsymbol{x}_2)\cdots p(\boldsymbol{x}_V|\boldsymbol{x}_{[V-1]})$$

- Assuming binary r.v., $p(X_V \mid X_{[V-1]})$ has 2^{V-1} parameters
- Total # parameters $\sum_{1 \leq i \leq V} 2^{i-1} = 2^V - 1$

CONDITIONAL INDEPENDENCE

- X and Y are conditionally independent given Z iff

$$p(X, Y | Z) = p(X | Z)p(Y | Z)$$

- Implies

$$p(X | Y, Z) = p(X, Y | Z)/p(Y | Z) = p(X | Z)$$

- Denoted

$$X \perp Y | Z$$

EX. WHERE IND. OBVIOUSLY FACILITATES

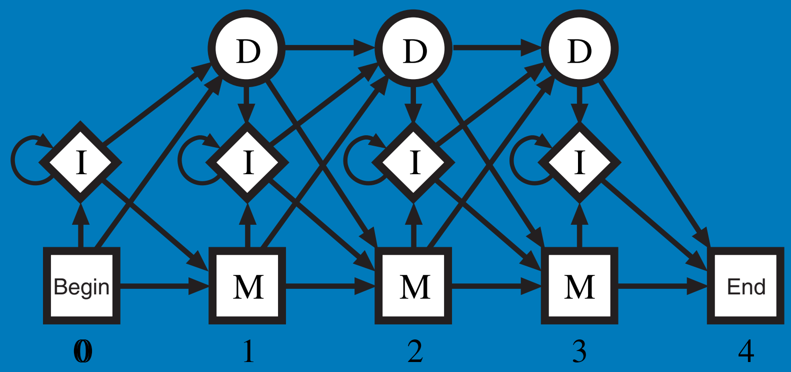
$$p(\underbrace{\boldsymbol{x}_{[V]}}_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_V}) = p(\boldsymbol{x}_1)p(\boldsymbol{x}_2|\boldsymbol{x}_1)p(\boldsymbol{x}_3|\boldsymbol{x}_1,\boldsymbol{x}_2)\cdots p(\boldsymbol{x}_V|\boldsymbol{x}_{[V-1]})$$

★Assume first order Markov property $\boldsymbol{x}_{t+1} \perp \boldsymbol{x}_{[t-1]}|\boldsymbol{x}_t$

i.e., if time ordered, future independent of past given present

★Then $$p(\boldsymbol{x}_{[V]}) = p(\boldsymbol{x}_1)\prod_{t=1}^{V-1} p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t)$$

SPECIAL CASE: HIDDEN MARKOV MODEL (HMM)



$$\begin{array}{ccccccc} z_1 & \longrightarrow & z_2 & \longrightarrow & z_3 & \cdots & z_V \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ x_1 & & x_2 & & x_3 & & x_V \end{array}$$

- Z_i hidden
- X_i observable
- Hidden often not observable when training, never when applying

SPECIAL CASE: HIDDEN MARKOV MODEL (HMM)

Combinations of the transition distributions

$$\begin{array}{ccccccccccc} Z_1 & \to & Z_2 & \to & Z_3 & \to & \cdots & \to & Z_T & \to & Z_{T+1} \\ \downarrow & & \downarrow & & \downarrow & & & & \downarrow & & \\ x_1 & & x_2 & & x_3 & & & & x_T & & \end{array}$$

Combinations of emission the emission distribution

All the same

State	1	2	3	4
1	A_{11}	A_{21}	A_{31}	A_{41}
2	A_{12}	A_{22}	A_{32}	A_{42}
3	A_{13}	A_{23}	A_{33}	A_{43}
4	A_{14}	A_{24}	A_{34}	A_{44}

$$
\begin{array}{ccccccccccc}
Z_1 & \rightarrow & Z_2 & \rightarrow & Z_3 & \rightarrow & \cdots & \rightarrow & Z_T & \rightarrow & Z_{T+1} \\
\downarrow & & \downarrow & & \downarrow & & & & \downarrow & & \\
x_1 & & x_2 & & x_3 & & & & x_T & &
\end{array}
$$

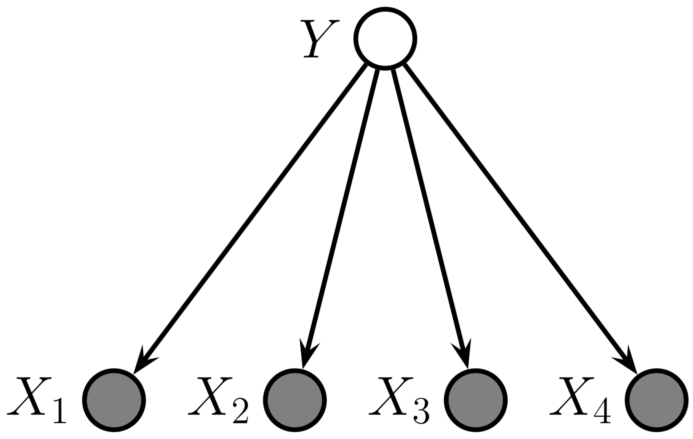
TRANSITION PROBABILITIES FOR 4 STATES HMM

EMISSION PROBABILITIES - HMM WITH 4 STATES & 3 SYMBOLS

$$
\begin{tikzcd}
Z_1 \arrow[r] \arrow[d] & Z_2 \arrow[r] \arrow[d] & Z_3 \arrow[r] \arrow[d] & \cdots \arrow[r] & Z_T \arrow[r] \arrow[d] & Z_{T+1} \\
x_1 & x_2 & x_3 & & x_T &
\end{tikzcd}
$$

All the same

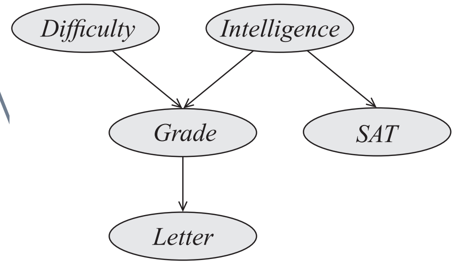
State\Symb	1	2	3
1	B_{11}	B_{21}	B_{31}
2	B_{12}	B_{22}	B_{32}
3	B_{13}	B_{23}	B_{33}
4	B_{14}	B_{24}	B_{34}



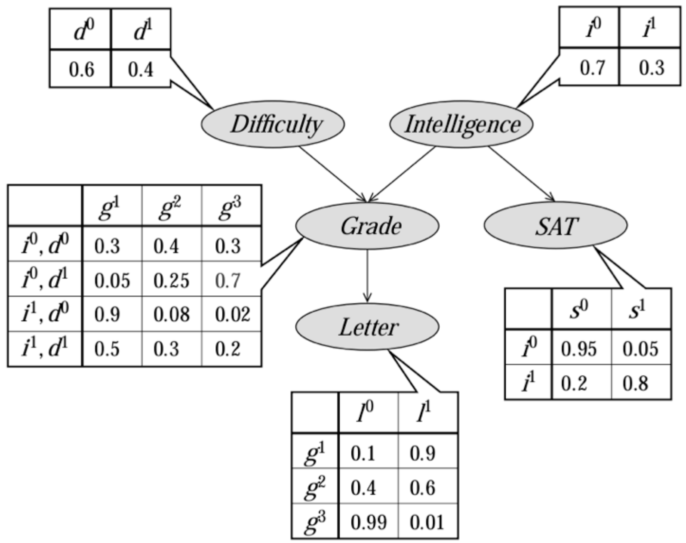
SPECIAL CASE: NAIVE BAYES CLASSIFIER

$$p(\boldsymbol{x}, y) = p(y) \prod_{t=1}^{4} p(x_t | y)$$

FACTORIZATION - A BINARY EXAMPLE



D	I	S	G	L
0	1	1	1	1
1	1	1	0	0
1	1	0	0	1
1	0	0	0	0
1	1	0	0	1



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

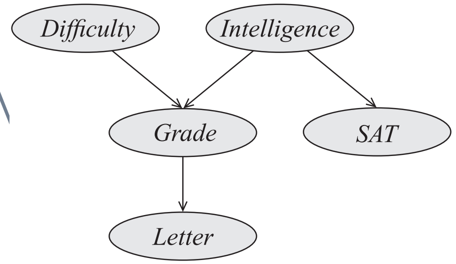
	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

Given data and GM with CPDs (new CPDs on a need to know basis)

FACTORIZATION - A EXAMPLE

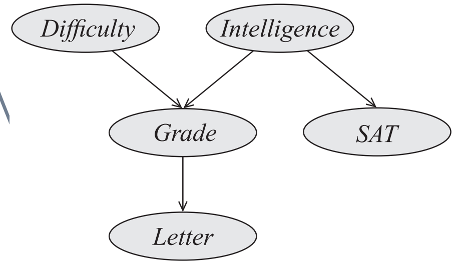


Difficulty (1 diff), Intelligence (1 int), Sat (1 good), Grade (1 good), Letter (1good)

D	I	S	G	L
0	1	1	1	1
1	1	1	0	0
1	1	0	0	1
1	0	0	0	0
1	1	0	0	1

$$
\begin{aligned}
L(\boldsymbol{\theta}; \mathcal{D}) = & \, p(0,1,1,1,1|\boldsymbol{\theta}) p(1,1,1,0,0|\boldsymbol{\theta}) \\
& \, p(1,1,0,0,1|\boldsymbol{\theta}) p(1,0,0,0,0|\boldsymbol{\theta}) \\
& \, p(1,1,0,0,1|\boldsymbol{\theta})
\end{aligned}
$$

FACTORIZATION - AN EXAMPLE



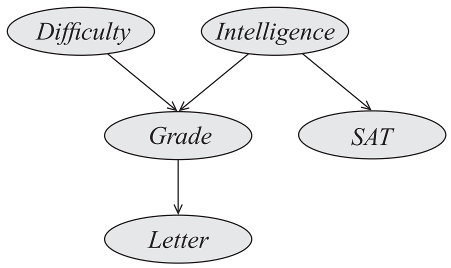
Difficulty (1 diff), Intelligence (1 int), Sat (1 good), Grade (1 good), Letter (1good)

D	I	S	G	L
0	1	1	1	1
1	1	1	0	0
1	1	0	0	1
1	0	0	0	0
1	1	0	0	1

$$
\begin{aligned}
L(\boldsymbol{\theta}; \mathcal{D}) = \; & p(0,1,1,1,1|\boldsymbol{\theta})p(1,1,1,0,0|\boldsymbol{\theta}) \\
& p(1,1,0,0,1|\boldsymbol{\theta})p(1,0,0,0,0|\boldsymbol{\theta}) \\
& p(1,1,0,0,1|\boldsymbol{\theta})
\end{aligned}
$$

$$
\begin{aligned}
L(\boldsymbol{\theta}; \mathcal{D}) = \; & p(D = (0,1,1,1,1)|\boldsymbol{\theta}_D) \\
& p(I = (1,1,1,0,1)|\boldsymbol{\theta}_I) \\
& p(S = (1,1,0,0,0)|I = (1,1,1,0,1), \boldsymbol{\theta}_S) \\
& p(G = (1,0,0,0,0)|D = (0,1,1,1,1), I = (1,1,1,0,1), \boldsymbol{\theta}_G) \\
& p(L = (1,0,1,0,1)|G = (1,0,0,0,0), \boldsymbol{\theta}_L)
\end{aligned}
$$

AN EXAMPLE



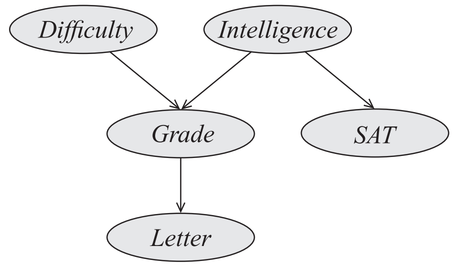
Difficulty (1 diff), Intelligence (1 int), Sat (1 good), Grade (1 good), Letter (1good)

D	I	S	G	L
0	1	1	1	1
1	1	1	0	0
1	1	0	0	1
1	0	0	0	0
1	1	0	0	1

θ_D	D=0	D=1
	2/5	3/5

$$
\begin{aligned}
L(\boldsymbol{\theta};\mathcal{D}) = {} & p(D=(0,1,1,1,1)|\boldsymbol{\theta}_D) \\
& p(I=(1,1,1,0,1)|\boldsymbol{\theta}_I) \\
& p(S=(1,1,0,0,0)|I=(1,1,1,0,1),\boldsymbol{\theta}_S) \\
& p(G=(1,0,0,0,0)|D=(0,1,1,1,1),I=(1,1,1,0,1),\boldsymbol{\theta}_G) \\
& p(L=(1,0,1,0,1)|G=(1,0,0,0,0),\boldsymbol{\theta}_L)
\end{aligned}
$$

AN EXAMPLE



Difficulty (1 diff), Intelligence (1 int), Sat (1 good), Grade (1 good), Letter (1good)

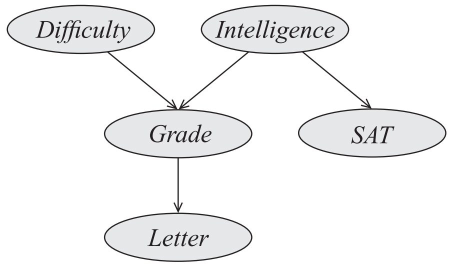
D	I	S	G	L
0	1	1	1	1
1	1	1	0	0
1	1	0	0	1
1	0	0	0	0
1	1	0	0	1

θ_D	D=0	D=1
	2/5	3/5

$$L(\boldsymbol{\theta}; \mathcal{D}) = \frac{2}{5}\left(\frac{3}{5}\right)^4$$

$$p(I = (1,1,1,0,1)|\boldsymbol{\theta}_I)$$
$$p(S = (1,1,0,0,0)|I = (1,1,1,0,1), \boldsymbol{\theta}_S)$$
$$p(G = (1,0,0,0,0)|D = (0,1,1,1,1), I = (1,1,1,0,1), \boldsymbol{\theta}_G)$$
$$p(L = (1,0,1,0,1)|G = (1,0,0,0,0), \boldsymbol{\theta}_L)$$

AN EXAMPLE



Difficulty (1 diff), Intelligence (1 int), Sat (1 good), Grade (1 good), Letter (1good)

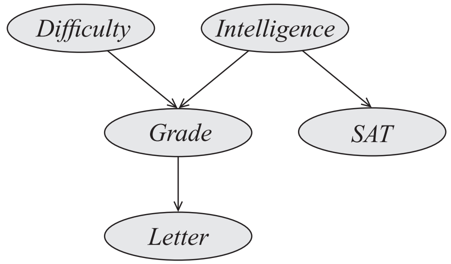
D	I	S	G	L
0	1	1	1	1
1	1	1	0	0
1	1	0	0	1
1	0	0	0	0
1	1	0	0	1

θ_I	I=0	I=1
	1/4	3/4

$$L(\boldsymbol{\theta}; \mathcal{D}) = \frac{2}{5}\left(\frac{3}{5}\right)^4$$

$$p(I = (1,1,1,0,1)|\boldsymbol{\theta}_I)$$
$$p(S = (1,1,0,0,0)|I = (1,1,1,0,1), \boldsymbol{\theta}_S)$$
$$p(G = (1,0,0,0,0)|D = (0,1,1,1,1), I = (1,1,1,0,1), \boldsymbol{\theta}_G)$$
$$p(L = (1,0,1,0,1)|G = (1,0,0,0,0), \boldsymbol{\theta}_L)$$

AN EXAMPLE



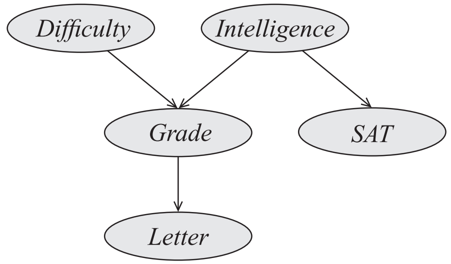
Difficulty (1 diff), Intelligence (1 int), Sat (1 good), Grade (1 good), Letter (1good)

D	I	S	G	L
0	1	1	1	1
1	1	1	0	0
1	1	0	0	1
1	0	0	0	0
1	1	0	0	1

θ_I	I=0	I=1
	1/4	3/4

$$L(\boldsymbol{\theta}; \mathcal{D}) = \frac{2}{5}\left(\frac{3}{5}\right)^4 \frac{1}{4}\left(\frac{3}{4}\right)^4$$
$$p(S = (1,1,0,0,0) | I = (1,1,1,0,1), \boldsymbol{\theta}_S)$$
$$p(G = (1,0,0,0,0) | D = (0,1,1,1,1), I = (1,1,1,0,1), \boldsymbol{\theta}_G)$$
$$p(L = (1,0,1,0,1) | G = (1,0,0,0,0), \boldsymbol{\theta}_L)$$

AN EXAMPLE



Difficulty (1 diff), Intelligence (1 int), Sat (1 good), Grade (1 good), Letter (1good)

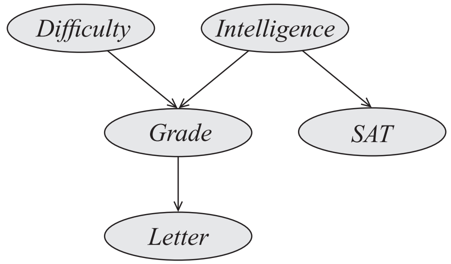
D	I	S	G	L
0	1	1	1	1
1	1	1	0	0
1	1	0	0	1
1	0	0	0	0
1	1	0	0	1

θ_S	S=0	S=1
I=0	1	0
I=1	1/6	5/6

$$L(\boldsymbol{\theta};\mathcal{D}) = \frac{2}{5}\left(\frac{3}{5}\right)^4 \frac{1}{4}\left(\frac{3}{4}\right)^4$$

$$p(S=(1,1,0,0,0)|I=(1,1,1,0,1),\boldsymbol{\theta}_S)$$
$$p(G=(1,0,0,0,0)|D=(0,1,1,1,1),I=(1,1,1,0,1),\boldsymbol{\theta}_G)$$
$$p(L=(1,0,1,0,1)|G=(1,0,0,0,0),\boldsymbol{\theta}_L)$$

AN EXAMPLE



Difficulty (1 diff), Intelligence (1 int), Sat (1 good), Grade (1 good), Letter (1good)

D	I	S	G	L
0	1	1	1	1
1	1	1	0	0
1	1	0	0	1
1	0	0	0	0
1	1	0	0	1

θ_S	S=0	S=1
I=0	1	0
I=1	1/6	5/6

$$L(\boldsymbol{\theta};\mathcal{D}) = \frac{2}{5}\left(\frac{3}{5}\right)^4 \frac{1}{4}\left(\frac{3}{4}\right)^4 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^2$$
$$p(G=(1,0,0,0,0)|D=(0,1,1,1,1), I=(1,1,1,0,1), \boldsymbol{\theta}_G)$$
$$p(L=(1,0,1,0,1)|G=(1,0,0,0,0), \boldsymbol{\theta}_L)$$

AN EXAMPLE

Difficulty (1 diff), Intelligence (1 int), Sat (1 good), Grade (1 good), Letter (1good)

D	I	S	G	L
0	1	1	1	1
1	1	1	0	0
1	1	0	0	1
1	0	0	0	0
1	1	0	0	1

	Less	Better
θ_G	G=0	G=1
D=0, I=0	1/2	1/2
D=1, I=0	3/5	2/5
D=0, I=1	1/10	9/10
D=1, I=1	2/5	3/5

$$L(\boldsymbol{\theta};\mathcal{D}) = \frac{2}{5}\left(\frac{3}{5}\right)^4\frac{1}{4}\left(\frac{3}{4}\right)^4\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^2$$

$$p(G=(1,0,0,0,0)|D=(0,1,1,1,1),I=(1,1,1,0,1),\boldsymbol{\theta}_G)$$

$$p(L=(1,0,1,0,1)|G=(1,0,0,0,0),\boldsymbol{\theta}_L)$$

AN EXAMPLE

Difficulty (1 diff), Intelligence (1 int), Sat (1 good), Grade (1 good), Letter (1good)

D	I	S	G	L
0	1	1	1	1
1	1	1	0	0
1	1	0	0	1
1	0	0	0	0
1	1	0	0	1

θ_G	G=0	G=1
D=0, I=0	1/2	1/2
D=1, I=0	3/5	2/5
D=0, I=1	1/10	9/10
D=1, I=1	2/5	3/5

$$L(\boldsymbol{\theta}; \mathcal{D}) = \frac{2}{5}\left(\frac{3}{5}\right)^4 \frac{1}{4}\left(\frac{3}{4}\right)^4 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^2 \frac{9}{10}\left(\frac{2}{5}\right)^3 \frac{3}{5}$$
$$p(L = (1,0,1,0,1) | G = (1,0,0,0,0), \boldsymbol{\theta}_L)$$

AN EXAMPLE

Difficulty (1 diff), Intelligence (1 int), Sat (1 good), Grade (1 good), Letter (1good)

D	I	S	G	L
0	1	1	1	1
1	1	1	0	0
1	1	0	0	1
1	0	0	0	0
1	1	0	0	1

θ_L	L=0	L=1
G=0	2/3	1/3
G=1	0	1

$$L(\boldsymbol{\theta};\mathcal{D}) = \frac{2}{5}\left(\frac{3}{5}\right)^4 \frac{1}{4}\left(\frac{3}{4}\right)^4 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^2 \frac{9}{10}\left(\frac{2}{5}\right)^3 \frac{3}{5}$$

$$p(L=(1,0,1,0,1)|G=(1,0,0,0,0),\boldsymbol{\theta}_L)$$

AN EXAMPLE

Difficulty (1 diff), Intelligence (1 int), Sat (1 good), Grade (1 good), Letter (1good)

D	I	S	G	L
0	1	1	1	1
1	1	1	0	0
1	1	0	0	1
1	0	0	0	0
1	1	0	0	1

θ_L	L=0	L=1
G=0	2/3	1/3
G=1	0	1

$$L(\boldsymbol{\theta};\mathcal{D}) = \frac{2}{5}\left(\frac{3}{5}\right)^4 \frac{1}{4}\left(\frac{3}{4}\right)^4 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^2 \frac{9}{10}\left(\frac{2}{5}\right)^3 \frac{3}{5}\left(\frac{2}{3}\right)^2 \left(\frac{1}{3}\right)^2$$

FACTORIZATION - AN EXAMPLE

"Row wise - per data point"

D	I	S	G	L
0	1	1	1	1
1	1	1	0	0
1	1	0	0	1
1	0	0	0	0
1	1	0	0	1

$$\begin{aligned} L(\boldsymbol{\theta}; \mathcal{D}) = \; & p(0,1,1,1,1|\boldsymbol{\theta}) p(1,1,1,0,0|\boldsymbol{\theta}) \\ & p(1,1,0,0,1|\boldsymbol{\theta}) p(1,0,0,0,0|\boldsymbol{\theta}) \\ & p(1,1,0,0,1|\boldsymbol{\theta}) \end{aligned}$$

"Column wise - per random variable - per family"

$$\begin{aligned} L(\boldsymbol{\theta}; \mathcal{D}) = \; & p(D = (0,1,1,1,1)|\boldsymbol{\theta}_D) \\ & p(I = (1,1,1,0,1)|\boldsymbol{\theta}_I) \\ & p(S = (1,1,0,0,0)|I = (1,1,1,0,1), \boldsymbol{\theta}_S) \\ & p(G = (1,0,0,0,0)|D = (0,1,1,1,1), I = (1,1,1,0,1), \boldsymbol{\theta}_G) \\ & p(L = (1,0,1,0,1)|G = (1,0,0,0,0), \boldsymbol{\theta}_L) \end{aligned}$$

CAN WE GET MLE?

Difficulty (1 diff), Intelligence (1 int), Sat (1 good), Grade (1 good), Letter (1good)

D	I	S	G	L
0	1	1	1	1
1	1	1	0	0
1	1	0	0	1
1	0	0	0	0
1	1	0	0	1

θ_L	L=0	L=1
G=0	2/3	1/3
G=1	0	1

$$L(\boldsymbol{\theta}; \mathcal{D}) = \frac{2}{5}\left(\frac{3}{5}\right)^4 \frac{1}{4}\left(\frac{3}{4}\right)^4 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^2 \frac{9}{10}\left(\frac{2}{5}\right)^3 \frac{3}{5}\left(\frac{2}{3}\right)^2 \left(\frac{1}{3}\right)^2$$

CAN WE GET MLE?

Difficulty (1 diff), Intelligence (1 int), Sat (1 good), Grade (1 good), Letter (1good)

D	I	S	G	L
0	1	1	1	1
1	1	1	0	0
1	1	0	0	1
1	0	0	0	0
1	1	0	0	1

θ_L	L=0	L=1
G=0	?	?
G=1	?	?

$$L(\boldsymbol{\theta}; \mathcal{D}) = \frac{2}{5}\left(\frac{3}{5}\right)^4 \frac{1}{4}\left(\frac{3}{4}\right)^4 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^2 \frac{9}{10}\left(\frac{2}{5}\right)^3 \frac{3}{5}\left(\frac{2}{3}\right)^2 \left(\frac{1}{3}\right)^2$$

CAN WE GET MLE?

Difficulty (1 diff), Intelligence (1 int), Sat (1 good), Grade (1 good), Letter (1good)

D	I	S	G	L
0	1	1	1	1
1	1	1	0	0
1	1	0	0	1
1	0	0	0	0
1	1	0	0	1

θ_L	L=0	L=1
G=0	1/2	1/2
G=1	0	1

MLE!

$$L(\boldsymbol{\theta}; \mathcal{D}) = \frac{2}{5}\left(\frac{3}{5}\right)^4 \frac{1}{4}\left(\frac{3}{4}\right)^4 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^2 \frac{9}{10}\left(\frac{2}{5}\right)^3 \frac{3}{5}\left(\frac{2}{3}\right)^2 \left(\frac{1}{3}\right)^2$$

CATEGORICAL – NOTATION

- ★ For a v ∈ [M],

  values $k \in S_v$

  Cartesian product

  combined values $c \in C_v = \prod_{s \in \mathrm{pa}(v)} S_s$

- ★ Cat CPDs

  where $P(x_v | x_{\mathrm{pa}(v)} = c) = \mathrm{Cat}(\boldsymbol{\theta}_{vc})$

  and $\theta_{vck} = P(X_v = k \,|\, X_{\mathrm{pa}(v)} = c)$

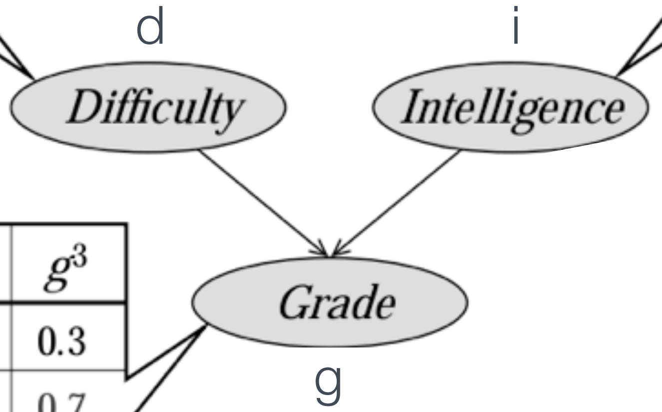
NOTATION EXAMPLE

$S_d = \{0,1\}$

$S_i = \{0,1\}$

d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3



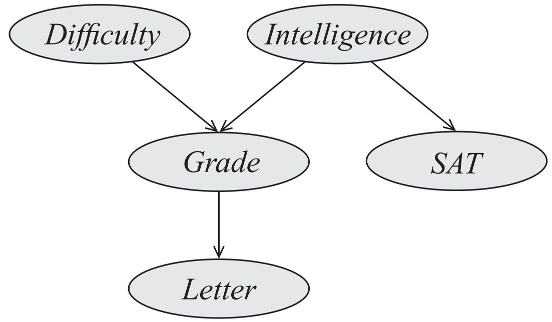
		g^1	g^2	g^3
$\mathrm{Cat}(\boldsymbol{\theta}_{g\langle 0,0\rangle})$	i^0, d^0	0.3	0.4	0.3
$\mathrm{Cat}(\boldsymbol{\theta}_{g\langle 0,1\rangle})$	i^0, d^1	0.05	0.25	0.7
$\mathrm{Cat}(\boldsymbol{\theta}_{g\langle 1,0\rangle})$	i^1, d^0	0.9	0.08	0.02
$\mathrm{Cat}(\boldsymbol{\theta}_{g\langle 1,1\rangle})$	i^1, d^1	0.5	0.3	0.2

$\theta_{g\langle 1,1\rangle 2}$

$S_g = \{1,2,3\}$

$$C_g = \prod_{s \in \mathrm{pa}(g)} S_s = S_i \times S_d = \{\langle 0,0\rangle, \langle 0,1\rangle, \langle 1,0\rangle, \langle 1,1\rangle\}$$

FACTORIZATION - AN EXAMPLE



$$
\begin{aligned}
L(\boldsymbol{\theta};\mathcal{D}) = \, & p(D=(0,1,1,1,1)|\boldsymbol{\theta}_D) \\
& p(I=(1,1,1,0,1)|\boldsymbol{\theta}_I) \\
& p(S=(1,1,0,0,0)|I=(1,1,1,0,1),\boldsymbol{\theta}_S) \\
& p(G=(1,0,0,0,0)|D=(0,1,1,1,1),I=(1,1,1,0,1),\boldsymbol{\theta}_G) \\
& p(L=(1,0,1,0,1)|G=(1,0,0,0,0),\boldsymbol{\theta}_L)
\end{aligned}
$$

"Column wise"

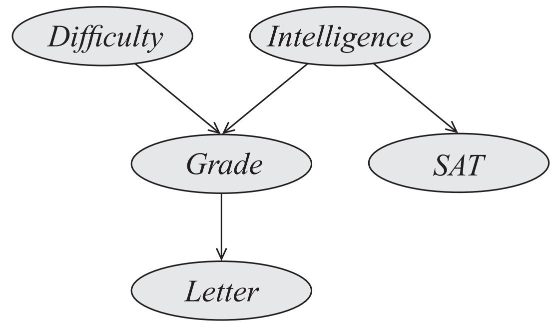
D	I	S	G	L
0	1	1	1	1
1	1	1	0	0
1	1	0	0	1
1	0	0	0	0
1	1	0	0	1

$$N_{vck} = \sum_{n=1}^{N} I(x_v^n = k, x_{pa(v)}^n = c)$$

$$N_{vc} = \sum_{n=1}^{N} I(x_{pa(v)}^n = c)$$

$$N_{G\langle 1,1\rangle 0} = \,?$$

FACTORIZATION - AN EXAMPLE



$$L(\boldsymbol{\theta}; \mathcal{D}) = p(D = (0,1,1,1,1)|\boldsymbol{\theta}_D)$$
$$p(I = (1,1,1,0,1)|\boldsymbol{\theta}_I)$$
$$p(S = (1,1,0,0,0)|I = (1,1,1,0,1), \boldsymbol{\theta}_S)$$
$$p(G = (1,0,0,0,0)|D = (0,1,1,1,1), I = (1,1,1,0,1), \boldsymbol{\theta}_G)$$
$$p(L = (1,0,1,0,1)|G = (1,0,0,0,0), \boldsymbol{\theta}_L)$$

"Column wise"

D	I	S	G	L
0	1	1	1	1
1	1	1	0	0
1	1	0	0	1
1	0	0	0	0
1	1	0	0	1

$$N_{vck} = \sum_{n=1}^{N} I(x_v^n = k, x_{pa(v)}^n = c)$$

$$N_{vc} = \sum_{n=1}^{N} I(x_{pa(v)}^n = c)$$

$$N_{G\langle 1,1\rangle 0} = 3$$

THE DGM LIKELIHOOD FACTORIZES

- Complete data

$$\mathscr{D} = \{x^1, \ldots, x^N\}$$

$$x^n = x_1^n, \ldots, x_V^n$$

- Likelihood

$$p(\mathscr{D} \mid \theta) = \prod_{n=1}^{N} p(x^n \mid \theta)$$

THE DGM LIKELIHOOD FACTORIZES

- Complete data

$$\mathscr{D} = \{x^1, \ldots, x^N\}$$

$$x^n = x_1^n, \ldots, x_V^n$$

- Likelihood

v's CPD

$$p(\mathscr{D} \mid \theta) = \prod_{n=1}^{N} p(x^n \mid \theta) = \prod_{n=1}^{N} \prod_{v=1}^{V} p(x_v^n \mid x_{\mathrm{pa}(v)}^n, \theta_v)$$

THE DGM LIKELIHOOD FACTORIZES

- Complete data

$$\mathcal{D} = \{x^1, \ldots, x^N\}$$

$$x^n = x_1^n, \ldots, x_V^n$$

- Likelihood

$$p(\mathcal{D} \mid \theta) = \prod_{n=1}^{N} p(x^n \mid \theta) = \prod_{n=1}^{N} \prod_{v=1}^{V} p(x_v^n \mid x_{\mathrm{pa}(v)}^n, \theta_v)$$

$$= \prod_{v=1}^{V} \prod_{n=1}^{N} p(x_v^n \mid x_{\mathrm{pa}(v)}^n, \theta_v)$$

THE DGM LIKELIHOOD FACTORIZES

- ★ Complete data

$$\mathscr{D} = \{x^1, \dots, x^N\}$$

$$x^n = x_1^n, \dots, x_V^n$$

- ★ Likelihood

$$p(\mathscr{D} \mid \theta) = \prod_{n=1}^{N} p(x^n \mid \theta) = \prod_{n=1}^{N} \prod_{v=1}^{V} p(x_v^n \mid x_{\mathrm{pa}(v)}^n, \theta_v)$$

$$= \prod_{v=1}^{V} \prod_{n=1}^{N} p(x_v^n \mid x_{\mathrm{pa}(v)}^n, \theta_v) = \prod_{v=1}^{V} p(\mathscr{D}_v \mid \mathscr{D}_{\mathrm{pa}(v)}, \theta_v)$$

values of v

values of v's parents

Called: decomposable likelihood (factorizes into family-factors)

PARAMETER AND COUNTS

- Complete data $\mathcal{D} = \{x^1, \ldots, x^N\}$

$$x^n = x_1^n, \ldots, x_V^n$$

- Likelihood

$$p(\mathcal{D} \mid \theta) = \prod_{n=1}^{N} p(x^n \mid \theta) = \prod_{n=1}^{N} \prod_{v=1}^{V} p(x_v^n \mid x_{\mathrm{pa}(v)}^n, \theta_v)$$

$$= \prod_{v=1}^{V} \prod_{n=1}^{N} p(x_v^n \mid x_{\mathrm{pa}(v)}^n, \theta_v) = \prod_{v=1}^{V} p(\mathcal{D}_v \mid \mathcal{D}_{\mathrm{pa}(v)}, \theta_v)$$

$$= \prod_{v=1}^{V} \prod_{c \in C_v} \prod_{k \in S_v} \theta_{vck}^{N_{vck}}$$

THE LIKELIHOOD FACTORIZES

- Complete data

$$\mathscr{D} = \{x^1, \ldots, x^N\}$$

$$x^n = x_1^n, \ldots, x_V^n$$

- Likelihood

$$p(\mathscr{D} \mid \theta) = \prod_{n=1}^{N} p(x^n \mid \theta) = \prod_{n=1}^{N} \prod_{v=1}^{V} p(x_v^n \mid x_{\mathrm{pa}(v)}^n, \theta_v)$$

$$= \prod_{v=1}^{V} \prod_{n=1}^{N} p(x_v^n \mid x_{\mathrm{pa}(v)}^n, \theta_v) = \prod_{v=1}^{V} p(\mathscr{D}_v \mid \mathscr{D}_{\mathrm{pa}(v)}, \theta_v)$$

$$= \prod_{v=1}^{V} \prod_{c \in C_v} \prod_{k \in S_v} \theta_{vck}^{N_{vck}}$$

So ML estimate

$$\theta_{vck} = N_{vck} / N_{vc}$$

PLATE NOTATION

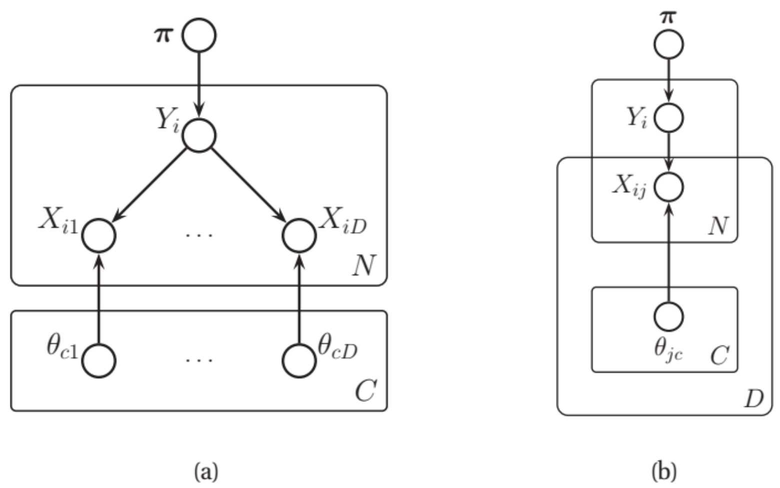
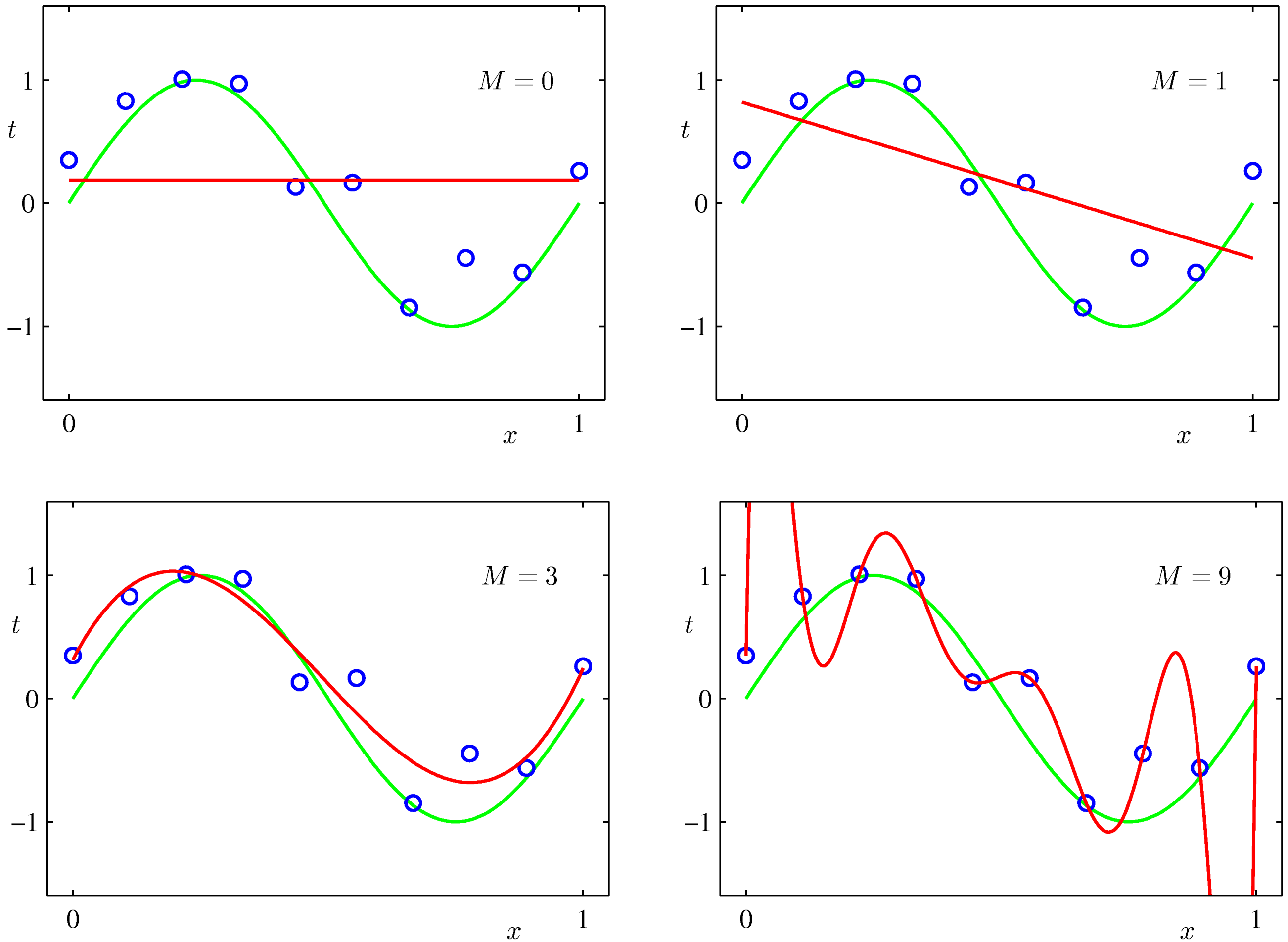


Figure 10.8 Naive Bayes classifier as a DGM. (a) With single plates. (b) WIth nested plates.

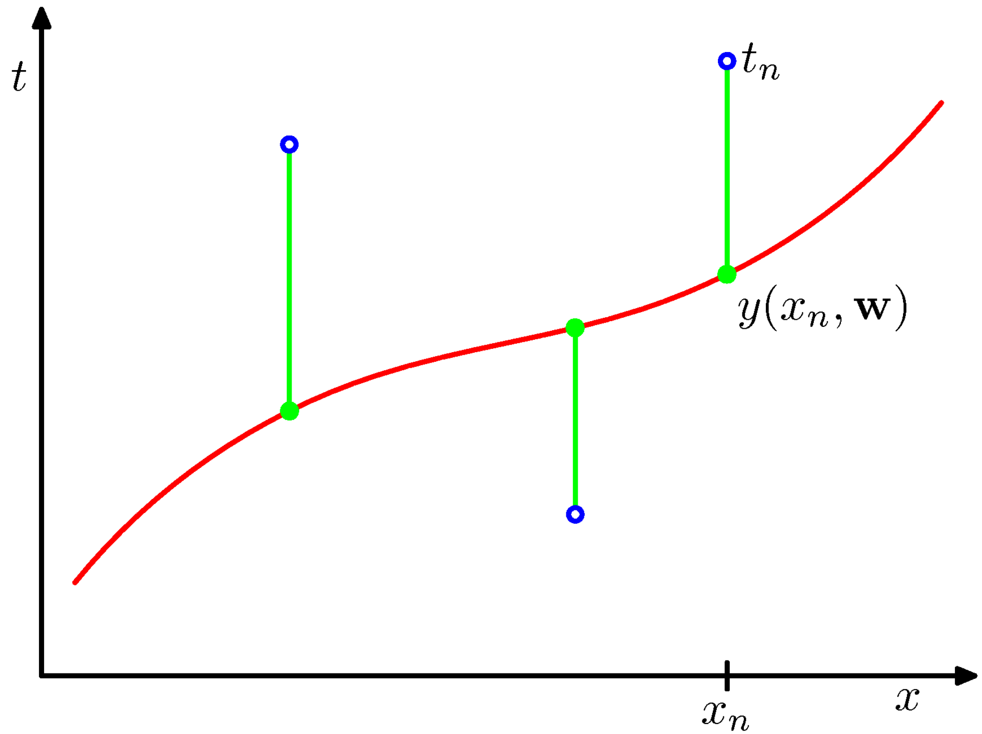
MODELS & PARAMETERS

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$



LIKELIHOOD

For N data points $$\prod_{n=1}^{N} \mathcal{N}(y(x_n, \mathbf{w}) - t_n | \mu, \sigma)$$



VISUALISING A MODEL- POLYNOMIAL REGRESSION

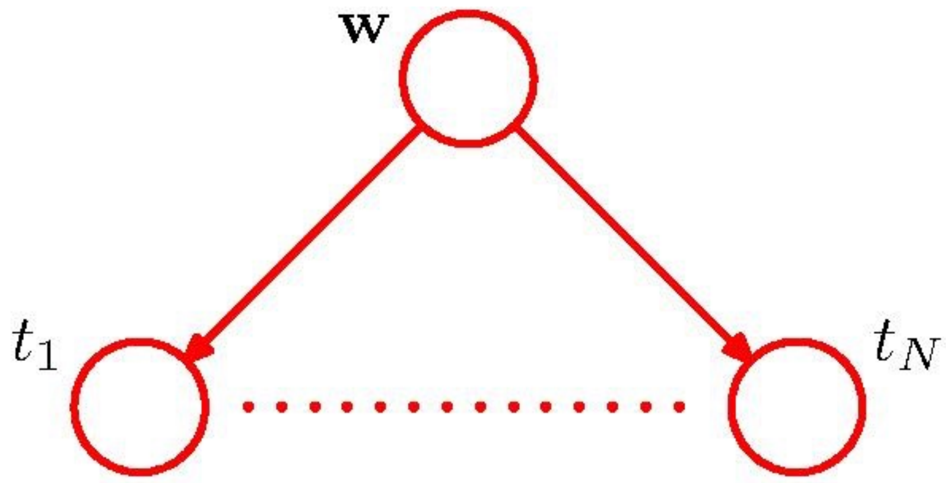
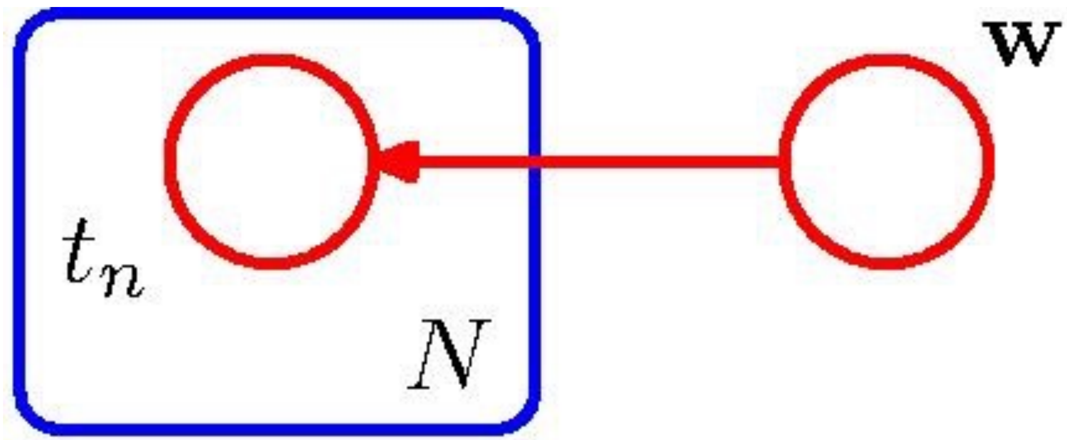
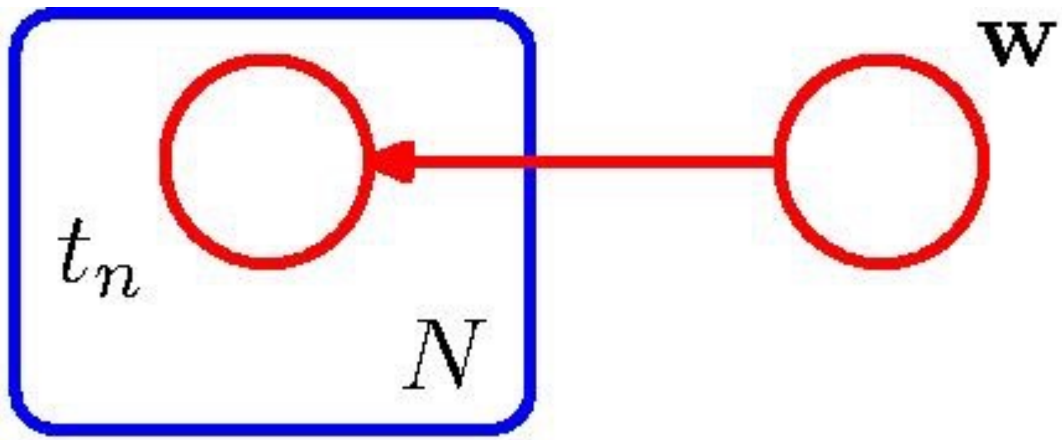


PLATE NOTATION



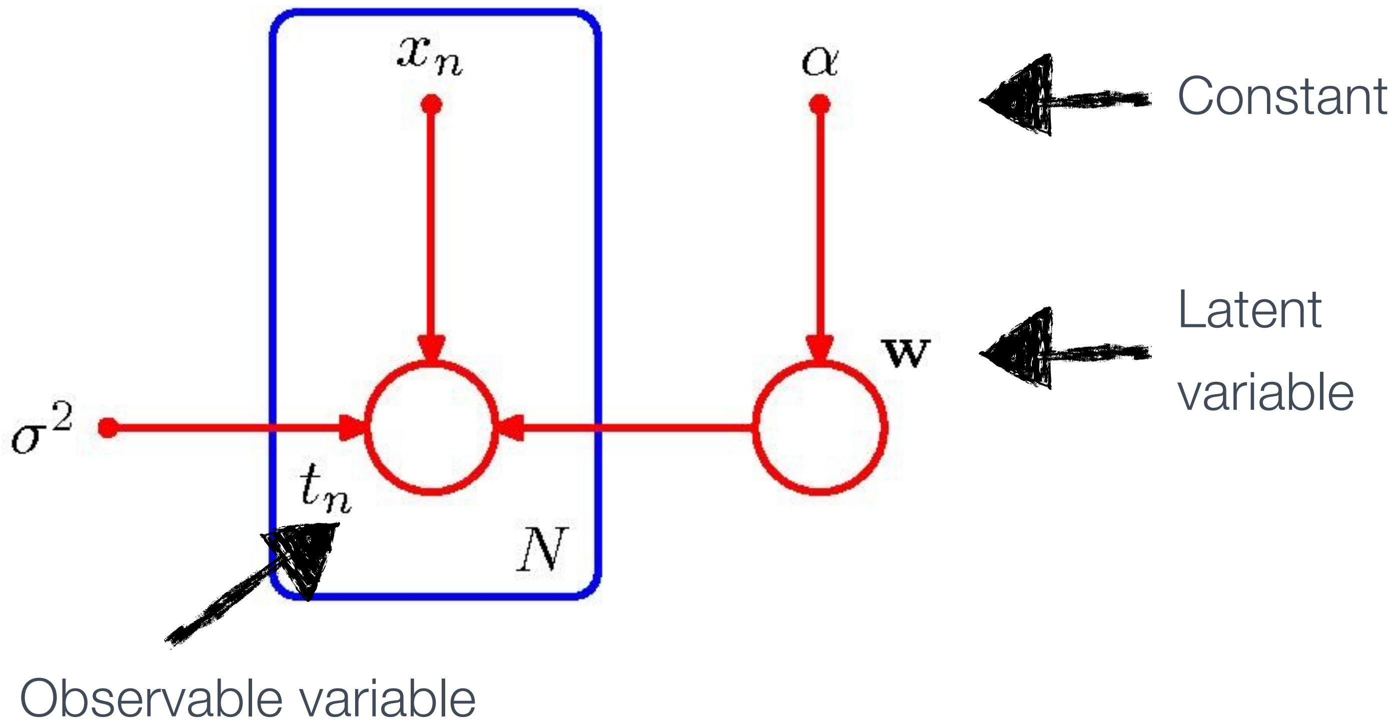
THE JOINT

$$p(\mathbf{t}, \mathbf{w}) = p(\mathbf{w}) \prod_{n=1}^{N} p(t_n|\mathbf{w})$$



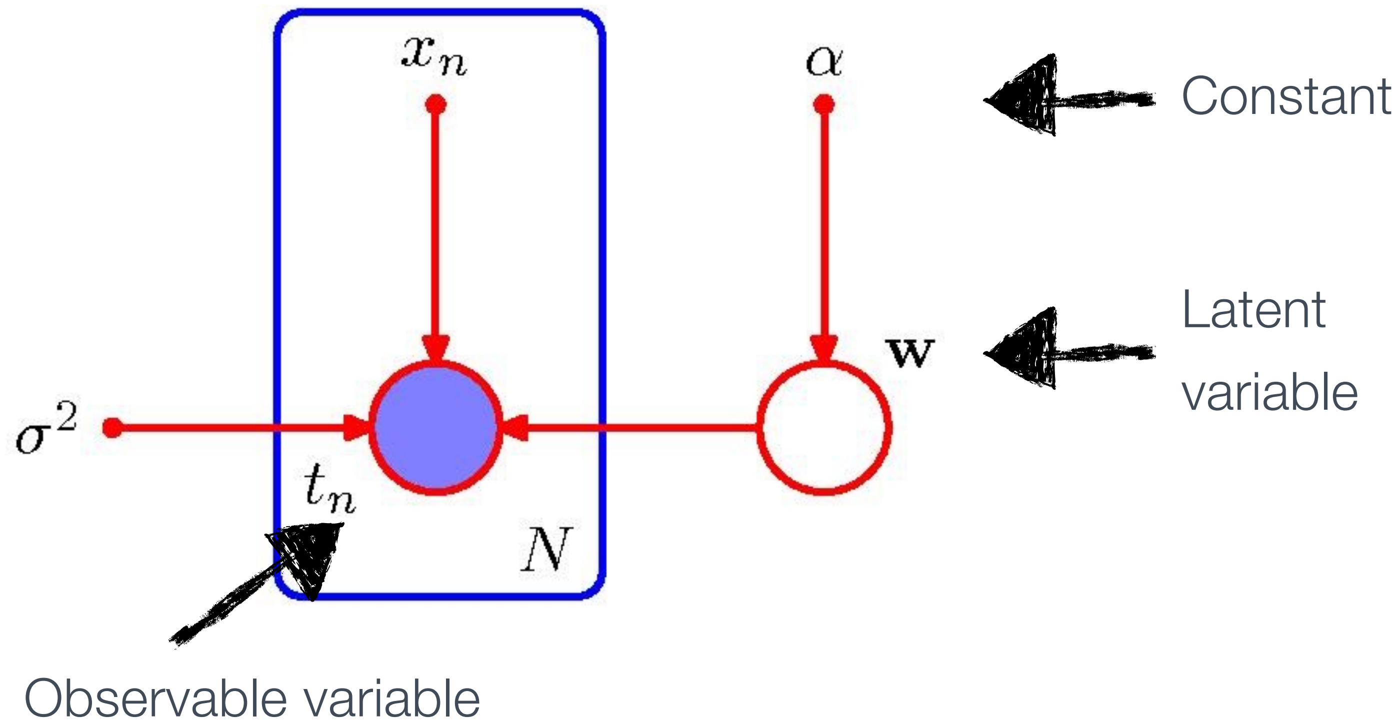
INCLUDING PARAMETERS

$$p(\mathbf{t}, \mathbf{w}|\mathbf{x}, \alpha, \sigma^2) = p(\mathbf{w}|\alpha) \prod_{n=1}^{N} p(t_n|\mathbf{w}, x_n, \sigma^2)$$



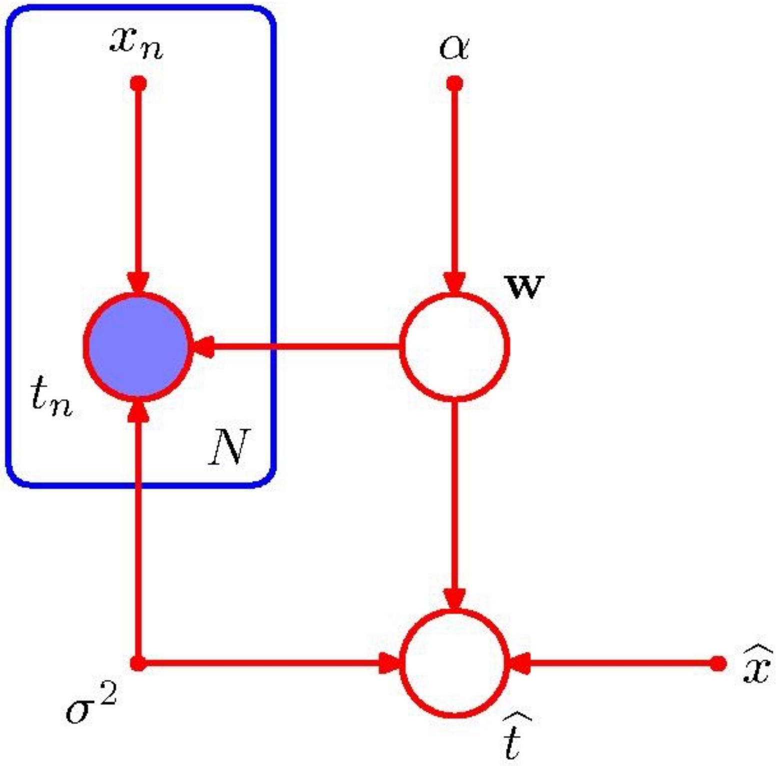
INCLUDING PARAMETERS

$$p(\mathbf{t}, \mathbf{w}|\mathbf{x}, \alpha, \sigma^2) = p(\mathbf{w}|\alpha) \prod_{n=1}^{N} p(t_n|\mathbf{w}, x_n, \sigma^2)$$



JOINT WITH NEW POINT

$$p(\widehat{t}, \mathbf{t}, \mathbf{w}|\widehat{x}, \mathbf{x}, \alpha, \sigma^2) = \left[\prod_{n=1}^{N} p(t_n|x_n, \mathbf{w}, \sigma^2)\right] p(\mathbf{w}|\alpha) p(\widehat{t}|\widehat{x}, \mathbf{w}, \sigma^2)$$



THE END